Deccan College

Building Centenary and Silver Jubilee Series

35

STUDIES IN HINDI-URDU

I

Introduction and Word Phonology

by

Ashok R. Kelkar

DECCAN COLLEGE
Postgraduate and Research Institute
Poona

STUDIES IN HINDI-URDU

I

Introduction and Word Phonology

by

Ashok Ramchandra Kelkar
M. A. (Poona), Ph. D. (Cornell)
Professor of Applied Linguistics, Centre of Advanced Study in Linguistics (University of Poona) at Deccan College

DECCAN COLLEGE
Postgraduate and Research Institute
Poona

1968

Code No.: BC 119

First edition : 1000 copies, 1968

xii + 87 pages + 1 foldout

Price Rs. 12.00

Printed by V. P. Bhagwat at Mouj Printing Bureau, Khatau Wadi, Bombay 4 and published by Dr. S. M. Katre of Deccan College Postgraduate and Research Institute, Yeravada, Poona 6

for

Dr. Sunitikumar Chatterji

who led my first steps into the wonderland of phonetics

FOREWORD

On the 15th of October 1964 the Deccan College celebrates the centenary of its main Building, and curiously enough this period coïncided with the Silver Jubilee of the Postgraduate and Research Institute which, as successor to the Deccan College, started functioning from 17th August 1939 when members of the teaching faculty reported on duty. When I suggested to members of our faculty the novel idea that the centenary should be celebrated by the publication of a hundred monographs representing the research carried on under the auspices of the Deccan College in its several departments they readily accepted the suggestion. These contributions are from present and past faculty members and research scholars of the Deccan College, giving a cross-section of the manifold research that it has sponsored during the past twentyfive years. From small beginnings in 1939 the Deccan College has now grown into a well developed and developing Research Institute and become a national centre in so far as Linguistics, Archaeology and Ancient Indian History, and Anthropology and Sociology are concerned. Its international status is attested by the location of the Indian Institute of German Studies (jointly sponsored by Deccan College and the Goethe Institute of Munich), the American Institute of Indian Studies and a branch of the Ecole Francaise d'Extreme-Orient in the campus of the Deccan College. The century of monographs not only symbolises the centenary of the original building and the silver jubilee of the Research Institute, but also the new spirit of critical enquiry and the promise of more to come.

7th March 1964 S. M. Katre

PREFACE

The aim of this and future studies in Hindī-Urdu is to provide a comprehensive framework for further field-work and discussion that could lead to a definitive description of contemporary standard Hindi-Urdu (or Hirdu for short). I am painfully aware that this presentation is based on insufficient fieldwork and insufficient tapping of secondary sources, i.e. previous descriptive and prescriptive studies. I have had no access to informants from Pakistan or to published work written in Urdu, for example. But repeated experience of frustrating discussions based on idiolect studies—especially in phonology—has convinced me that some framework is better than none, and that the whole gamut of linguistic systems that could possibly come under the rubric of contemporary standard Hindi or Urdu needs to be studied together.

In Figures 3 and 4 and the discussion pertaining to them I am certainly sticking my neck out, but a fresh statement was overdue for some years even if it then comes in for correction and further revision by more competent hands.

The arrangement and the style of presentation is intended to be in consonance with the essentially exploratory and preparatory nature of this study. This involves a certain repetitive and loose-knit quality, for which I crave the methodological purist's indulgence. The reader will notice that I have resorted to the practice of giving minimally and sub-minimally contrasting pairs for phonemic and morphophonemic distinctions. This practice is rather out of fashion now, but deserves to be revived.

It is difficult to write about Hindi-Urdu without exacerbating sensibilities. I may simply plead here that

my motive has only been the uncovering, recording, and systematizing of facts. If I have said something that either pleases or annoys anyone, that is a wholly unintended result. I have of course recorded educated speech in its unguarded moments too and duly labelled such alternants appropriately. Again, showing either approval or disapproval of these is furthest from my intentions.

Substantial portions of the phonological part (ch. 1) were presented in an earlier version at the Winter Seminar of Linguistics, Kurukshetra University, December 1963. I have benefited from the comments of my co-participants. I have pestered numerous colleagues and students at Agra (1958-62), Poona, and elsewhere for data, some of it semi-cooked. I must specially mention Dr. Murari Lal Upraiti (Agra University), Dr. Amba Prasad Suman (Dakshin Bharat Hindi Prachar Sabha, Madras), Dr. Gyan Chand Jain (Hamidia College, Bhopal), and Mr. Ram Prakash Dixit (University of California, Berkeley). Indeed, in the last two cases the debt comes near to being collaboration at some points. My views, however, often differ from theirs—which may be ascertained respectively from: Dr. JAIN's *Lisani mutalae* ("Linguistic studies"), Karachi: Anjuman Taraqqi-e-Urdu, 1964 and Mr. DIXIT's The Segmental phonemes of Contemporary Hindi, University of Texas M.A. dissertation, 1963. Dr. A. M. Ghatage (University of Poona and Deccan College), Dr. M. A. Mehendale (Deccan College), Dr. P. B. Pandit (University of Delhi), Dr. Masud Husain Khan (Osmania University, Hyderabad, Andhra Pradesh), Dr. Devishankar Dwivedi (Vikram University, Ujjain, M. P.), Mr. Stuart McGregor (Cambridge University), and Mr. Shri Krishna Swamy were kind enough to go through an earlier version of the completed manuscript and make useful comments. I have pleasure in expressing my sense of gratitude to all of these and to my predecessors in the fields of phonology

in general and of Hindi-Urdu phonology in particular. More specific acknowledgements have been made in the footnotes.

I deem it a privilege that this study is being associated, through the kindness of the authorities of Deccan College, with our Building Centenary and Silver Jubilee. Dr. S. M. Katre, Director, and Dr. N. G. Kalelkar, who looks after publications, have been especially solicitous. The manuscript was submitted in essentially the present form in April 1965. I have taken advantage of the interval before publication to make a few minor alterations and corrections. The Mouj Press have been most kind to me and to the manuscript.

I should like to invite readers to be kind enough to draw my attention to any errors of commission and omission and to the possibility of better and more varied illustrations. For shortcomings in these respects naturally I am alone responsible. The reader's attention is drawn to the table of Addenda and corrigenda.

22 April 1968 A. R. K.

CONTENTS

FIGURES

In addition there are seven Tables in the text (*on pp.* 22, 29, 45, 56, 58, 79, 82).

NOTE : *In Figure 4, columns 4 and 5 for* Mahajan and Nagari script *read* Mahājanī and Nāgarī scripts.

Chapter 0

INTRODUCTION

0.0. The Problem stated.

Descriptive linguists, in their justifiable anxiety to have a relatively homogeneous body of parole on which to base their study of the language system, tend to rely on idiolects as starting points. Eventually, however, the starting point has to be left behind. An idiolect is after all a point of reference that is extrinsic to the symbolic culture system as such : an idiolect is a biographical accident. With many other languages this does not involve a serious practical limitation. Contemporary standard Hindi-Urdu, however, consists of a gamut of intergraded variations that need to be studied together—within a single framework,

This is particularly true of phonology. Here the tolerance of the listeners for variety is very great. Purists who will haggle over points of grammar and vocabulary are cheerfully unaware of the wide variety of pronunciations disguised by uniformities of orthography.[1] The study by Gumperz and Naim,[2] which refuses to concentrate on some single variety and treat

1. In this detail, the Hindi-Urdu speech community resembles the French one. A story like that of Bernard Shaw's *Pygmalion* centred around RP is probably unthinkable in any country except England—and certainly not in France or Upper India. Exceptional cases may be noted, however, of intense awareness about the choice of phonemes—Urdu speakers about consonant phonemes in Persian loanwords (§ 1.2.1, especially 3–11), and Urdu and Hindi speakers about retroflex flaps and nasals (§ 1.2.3).

2. John J. Gumperz and C. M. Naim, Formal and informal standards in the Hindi regional language area, in *Linguistic diversity in South Asia...*, ed. C. A. Ferguson and John J. Gumperz, *IJAL* suppl., 1960 pp. 92-118.

it as a separate unit and takes instead all of them together as the unit of study, is therefore a welcome departure. It is, however, only a departure and needs to be followed up. Before we can go into the methodological preliminaries, it is necessary to go over some familiar ground and unsnarl the tangle of labels like Hindī, Urdū, Khaṛī Bolī, and the rest.

0. 1. Hindi-Urdu in its contemporary social-and-geographical setting.

If we leave out the Gypsy dialects, Afghoni, Sinhalese, and the Dardic-Kashmiri group, modern Indo-Aryan (or Indic) dialects probably form a dialect continuum (Figure 1) that displays variations from village to village and from caste-group to caste-group and that exhibits one notable peculiarity. It reenters upon itself, and this results in linguistic discontinuities, which suddenly bring home to us, as it were, the amount of diversification that the Indo-Aryan subfamily has undergone and the complicated nature of the history of this diversification. This total continuum is farmed out, so to say, among some selected koines that have received the stamp of state recognition and contemporary printed literature. These are Hindi-Urdu, Panjabi, Sindhi, Gujarati, Marathi, Oriya, Bengali, Assamese, and Nepali. Barring Hindi-Urdu, each of the rest (as also Kashmiri) has a clearly defined homeland—the area over which the koine dominates coincides with the geographical extension of the dialect-cluster which is the linguistic matrix of that koine. If we say that Bengali, Gujarati, Tamil, and the rest are the regionally "marked" koines of South Asia (to use the convenient cover term for India, Pakistan, Nepal, Bhutan, and Ceylon), Hindi-Urdu (or Hirdu as I propose to call it from now on) is, much like its predecessor Sanskrit and its contemporary rival English, the regionally "unmarked" koine of South Asia. When its homeland is defined, it is

done negatively as the residual area left by the regionally "marked" koines. This residual area is sometimes called Hindustan or Upper India or, with some inexactitude, North India. When we need some handy ethnic name for the people of this area, further excisions are made—namely, the Rajasthanis, the Panjabis (Lahnda and Panjabi speakers), possibly the Biharis (the state of Bihar includes Magahi and Maithili areas and a part of the Bhojpuri area), the Paharis (as a vague rubric for Indo-Aryan and Tibeto-Burman speaking people of the Himalayan and sub-Himalayan areas of Panjab and Uttar Pradesh), and of course the Tribal or Adivasi groups (in the Malvi, Bundeli, Bagheli, Chhattisgarhi, and Magahi areas). For the remainder there is no obvious name. The very artificiality of the terms U.P. (whether the old United Provinces of Agra and Oudh or the new Uttar Pradesh, the "northern region") and Madhya Pradesh (the "central region"—a descendant not of the ancient Madhya Desh but the British Central Provinces and Central India Agency) is revealing.

Historically the linguistic matrix of Hirdu (see Fig. 2 and the terminological note thereto) is the so-called Khaṛī Bolī dialect area (GRIERSON's "vernacular Hindustani" of Meerut, Bijnaur, Sahranpur straddling U. P., Haryana, and Delhi states in India). At the "folk" or low-brow or "little tradition" level[3] (*gãvārū bolī* or *ghar kī bolī*) today the dialects genetically closest to and culturally dominated by Hirdu are the Regional Hindustani of this area and Regional Dakhani spoken by Muslims in Maharashtra, Mysore, Andhra Pradesh (but not in Gujarat, Kerala, Madras, or Bengal) (Fig. 2, no. 5). More distantly related to but culturally dominated by Hirdu are the dialect-areas of Lahnda, Pahari, Bangaru, Rajasthani,

3. The terms "little tradition" and "great tradition" are of course to be taken in the anthropologist REDFIELD's sense.

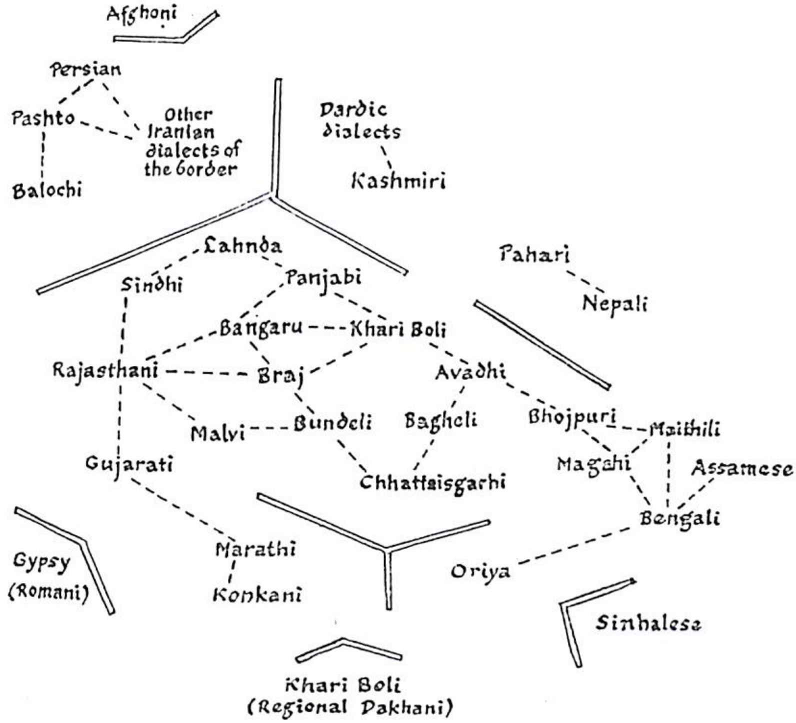


Figure 1. Contemporary spread of Indo-Aryan (Indic)

A double line indicates a present-day break between genetically related languages. A broken line indicates a present-day continuity between linguistically inter-shading and geographically contiguous dialects and groups of dialects. The distances of course have no scale. Khari Boli includes Hindi-Urdu, Regional Hindustani, and Regional Dakhani of the present day.

Malvi, Braj, Bundeli, Avadhi, Bagheli, Chhattisgarhi, Bhojpuri, Magahi, and Maithili (Fig. 2, no. 6; see also Fig. 1). There is oral "folk literature" in each of the regional speeches mentioned in this area (Fig. 2, nos. 5B, 6B). The language of this folk literature occasionally exhibits Hirdu interference (Fig. 2, Nos. 5A, 6A).

Hirdu can claim some as "adherent speakers" from among the speakers of all the dialects mentioned in the preceding paragraph, and their total number is large. But the "native speakers" of Hirdu in the strict linguistic sense for whom it is the language casually and informally learned as the first thing in childhood are considerably fewer and are limited to regionally uprooted families (labourers, bābū class, artisans, but not apparently Hindu traders) in the cities both in the "adherent area" (Delhi, Aligarh, Agra, Lucknow, Allahabad, Patna, Varanasi, Bhopal, Jabalpur, Lahore, Rawalpindi) and outside it (Bombay, Calcutta, Nagpur, Hyderabad (Andhra Pradesh), Karachi). (In the latter case they are emigrants from U.P., M.P., Bihar, and Haryana or from the Regional Dakhani speaking communities of the South.) What is special about Hirdu is that both "native" and "adherent" speakers have to be counted as equally "good informants" for eliciting texts and secondary responses (to questions such as "is this good Hindi/Urdu?"). Both types of speakers equally feel at home in Hirdu— enough so as not to worry about committing "mistakes" in the use of it, so as to be free from any sense of insecurity. The relations between Hirdu and the native speech varieties of the adherent speakers exhibit some of the familiar characteristics of what one may call "vertical diglossia",[4] although the "upper language" has "native speakers"

4. What I have called "vertical diglossia" has simply been called diglossia (Fr. diglossie). C. A. FERGUSON defines it (Diglossia, *Word* 15 (1959) 325-40, p. 336) as "a relatively stable linguistic situation in which, in addition to the

in this case and its linguistic matrix is limited only to one section of the "lower languages". Interference from the local "lower language" (Fig. 2, nos. 3A, 4A) is naturally expected in the casual use of Hirdu by "adherent speakers", but is also to be found to a lesser extent with "native speakers" of Hirdu (who may have picked it from "adherent speakers" or from using the local "lower language" as a second language).

The net result is that as a linguistic system Hirdu has no marked dialect variations; but it has the full gamut of styles (respectively nos. 1, 2, 3, and 4 in Fig. 2): *formalized highbrow* (poetry, learned discourse, oratory, religious sermons and the like in the "great tradition"[3] of urban centres of power, commerce, and religion), *formalized middlebrow* (popular printed literature, songs, and films; bhajan; the theatre has very little life; mass propaganda), *casual middlebrow* (everyday educated talk especially in linguistically mixed groups and within the regionally up-rooted upper or middle class family; private letter writing and newspapers waver between this and the previous styles; out of the four styles this is the most receptive to borrowings from English), and *casual lowbrow* (this is definitely substandard and outside the

---

primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature,...which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation". The examples he cites are classical and colloquial Arabic, standard and Swiss German, French and Haitian creole, classical and demotic Greek, classical and colloquial Chinese, sādhu and calita Bengali, and Tamil. (Telugu is probably also a case in point.) FERGUSON's paper and also Trevor HILL's (Institutional linguistics, *Orbis* 7 (1958) 441-55) helped me to clarify my ideas on the subject.

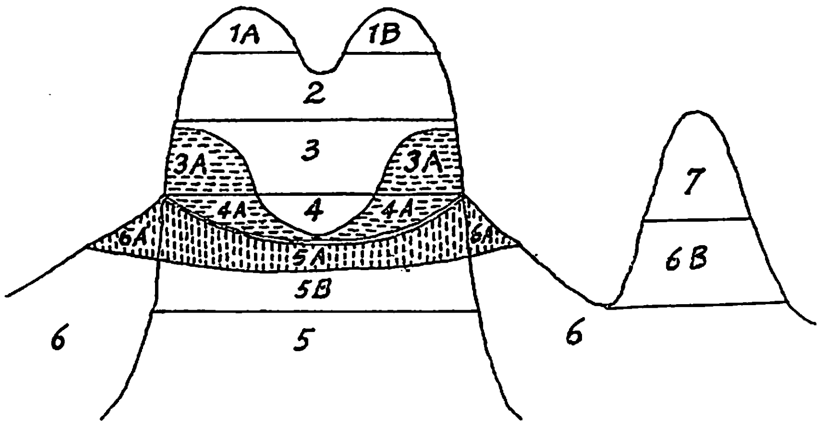


Figure 2. Hindi-Urdu in socio-linguistic perspective.

1 (i. e. 1A + 1B) Highbrow formalized Hirdu distinguishable into 1A Hindi and 1B Urdu. 2 Middlebrow formalized Hirdu. 3, 3A Middlebrow casual Hirdu. 1, 2, 3, 3A Hirdu in the Great Tradition. 4, 4A Lowbrow casual Hirdu. 5, 5A, 5B Regional Hindustani and Regional Dakhani. 6, 6A, 6B, 7 Other regional language-systems.

3A, 4A Casual Hirdu with regional colour. 5A, 6A Hirdu colour in regional dialects. 5B, 6B Formalized discourse in the Little or " Folk " Tradition. 7 Attempts to cultivate formalized discourse at the educated level.

While Khari Boli may be used as a comprehensive term to cover all the speech forms from 1 to 5 (i. e. 1A, 1B, 2, 3, 3A, 4, 4A, 5, 5A, 5B), Hindi-Urdu or Hirdu may be restricted to non-regional Khari Boli (i. e. 1A, 1B, 2, 3, 3A, 4, 4A) to the exclusion of Regional Hindustani and Regional Dakhani (i. e. excluding 5, 5A, 5B).

"great tradition"; everyday talk in lower-class, uneducated, urban milieu; this style, often called *bāzārū hindustānī*, is sometimes resorted to even by educated speakers and even in printed literature destined for the uneducated lower classes). All other variations can either be dismissed as occasional "regionalisms" (interference from non-Hirdu speeches) whose distribution varies from idiolect to idiolect or be placed in a horizontal polarization between "Hindi" and "Urdu" which reaches its maximum in the formalized highbrow style (Fig. 2, nos. 1A and 1B).[5] This polarization may be termed "horizontal diglossia". Hindi is associated with the Devanāgarī script (called Nāgarī for short) and the drawing upon Sanskrit for "higher vocabulary" and metrics, with secular nationalism and Hindu revivalism, and with what anthropologists have called "Sanskritization". It is the official language of the states of Rajasthan, Delhi, Himachal Pradesh, Uttar Pradesh, Madhya Pradesh, Bihar, Delhi, and Haryana (formerly a part of (East) Panjab); it is the union language of India. The cities of Varanasi, Allahabad, and Patna are centres of learning for Hindi. Urdu is associated with a modified form of the Perso-Arabic script and the drawing upon classical Persian (and through it, upon classical Arabic and upon Turkish) for "higher vocabulary" and metrics and with Muslim renascence and the courts of the Muslim princes. It is the official language of Jammu and Kashmir and of West Pakistan; along with Bengali it is one of the union

---

5. A convenient collection of varied sample texts in Devanagari orthography is available in Dhirendra VARMA (comp.), *Grāmīṇa hindī*, Allahabad: Sāhitya Bhavan, 8th printing, 1959, Appendix, pp. 63–70. With reference to fig. 2, "standard Hindustani" may be defined as 2–3, "Khari Boli" as 1–5 (i. e. inclusive of Regional Hindustani and Regional Dakhani), and "contemporary standard Hindi-Urdu" as 1–3. The term "Nagari" is best limited to the script and not applied to Hindi (fig. 2, no. 1A) as is sometimes done.

languages of Pakistan; it is one of the 15 languages in the Language Schedule of the Constitution of India. The cities of Delhi, Lucknow, Aligarh, Lahore, Hyderabad (Andhra Pradesh) (and, formerly, Agra) are centres of learning for Urdu. Attempts to set up a middle path—using the name Hindustānī and using the formalized middlebrow style—associated with Gandhian attempts at Hindi-Muslim unity can now be definitely relegated to history.[6] Outside the Hirdu adherent area, Hindi is cultivated by Hindu Rajasthanis in Maharashtra, in Calcutta, and elsewhere; by some Hindus in the Panjabi dialect-area of (East) Panjab; and by some Hindus (especially women) in Kashmir, while Urdu is cultivated by speakers of Panjabi, Kashmiri, and Regional Dakhani and by some Parsis in Bombay. Within the Hirdu adherent area of India, the cultivation of Urdu has been associated with Muslims, and with Kāyastha Hindus and immigrant Kashmiri Hindus and the cultivation of Hindi has been associated with Hindu women (even among Kayasthas and Kashmiris), Hindu baniyās, and paṇḍits (but not, apparently, with sādhūs).

The relations between Hirdu and the remaining Indo-Aryan speeches of South Asia today can now be set out with relative ease. There are five possibilities (regional, substandard speech and formalization of that speech in folk literature is common to all; figure 2, nos. 5, 6 and 5B, 6B) :

(1) 3A, 4A, and 5A (or 6A) also exist for Lahnda, Pahari, Bangaru, Malvi, Bundeli, Avadhi, Bagheli,

---

6. Recent attempts to publish Urdu texts in Devanagari editions is another attempt at a rapproachement. A strikingly similar example of horizontal diglossia is probably that of Serbo-Croat, Serb written in the Roman script and associated with the Roman Catholic Church and Croat written in the Cyrillic script and associated with the Greek Orthodox Church.

Chhattisgarhi, Bhojpuri and Magahi; Regional Hindustani and Regional Dakhani. (Thus for Lahnda, there is Hirdu-coloured Lahnda as well as Lahnda-coloured Hirdu).

(2) 3A, 4A, 6A, and also 7 exist for Braj, Maithili, and Rajasthani (especially of Jodhpur). (Thus, not only do we have Hirdu-coloured Braj or Braj-coloured Hirdu, but we have modern literary Braj also.)

(3) 3A, 4A, and also 7 (cultivated literature under the shadow of Hirdu) exist for Panjabi, Nepali, Sindhi,[7] Kashmiri. (Thus, for Panjabi there is Panjabi-coloured Hirdu and modern literary Panjabi but no Hirdu-coloured Panjabi.)

(4) 4A (but no 3A) and a full-fledged viable literature in the great tradition (and different styles corresponding to 1, 2, 3, 4 of Hirdu) exist for Gujarati, Marathi, Bengali, and Oriya. (The 4A of Bombay with its Marathi and Gujarati colour is institutionalized as *bambaiyā hindī* and has been much satirized. Note that Gujarati-coloured Hirdu, etc. do not exist at the middlebrow level.)

(5) Like the preceding, except that no 4A exists, is the case with Assamese and Sinhalese (which, thus, are on the same footing as the four cultivated Dravidian languages of South India.)

The first two groups constitute the "lower languages" of the Hirdu-based vertical diglossia. The third group is on the fringe. The much publicized "language problems" of India and Pakistan concern the relation between Hindi (or Urdu, respectively) and the languages of the last two groups.

---

7. Sindhi in Sind, West Pakistan is definitely under the shadow of Urdu; Sindhi as spoken by the Partition refugees in India is now more under the influence of Hindi, but is likely to have a brighter future.

0.2. Hindi-Urdu in its historical linguistic-and-literary setting.

The rise of Hindi and Urdu has often been recounted, and the account will not be repeated here.[8] Both are an outgrowth of Khaṛī Bolī, a branch of "Madhyadeshi" ("Western Hindi"), itself a branch of the "inner Central" group of the Indo-Aryan (Indic) sub-family of the Indo-European family (Fig. 3). Borrowings from classical Sanskrit into Hindi-Urdu (as indeed into all Modern Indo-Aryan languages) can be regarded in some sense as revivals. Borrowings from Panjabi, Braj, Kanauji, Avadhi into Khari Bolī, especially Hindi-Urdu, are in the nature of a diffused intimate interference and cannot always be distinguished from inter-grading isoglosses—on the whole Urdu tends to the West, toward Panjabi and Hindi tends to the East, toward Avadhi. Borrowings from Rajasthani and other "lower lauguages" are in the nature of substratum interference. Borrowings from Bengali, Tamil, and other koines of South Asia are in the nature of occasional cultural interference. Items seen as revivals from Sanskrit in Hirdu have often come via Bengali. Items traceable to Tamil have sometimes come via English. Finally, alongside of classical Sanskrit, classical Persian (including of course its Arabic borrowings) and English have been tapped wholesale as reservoirs for borrowing of "higher vocabulary". Sanskrit, Persian, and Arabic elements have also been used for new coinages and for translation loans from English; they are felt to be more "indigenous" than English ones.[9] Borrowings from Arabic and Turkish

8. See, for example, S. K. CHATTERJI, *Indo-Aryan and Hindi*, 2nd rev. ed., Calcutta: Mukhopadhyay, 1960, Part II, Lectures II and III; and especially A. BARANNIKOV, Modern literary Hindī, *BSOS* 8 (1935-7) : 2-3 (1936, Grierson number) 373-90.

9. These neo-Sanskrit, neo-Persian and, neo-Arabic elements are not borrowings in the strict sense; they are more like the neo-Latin elements in modern European languages. Historically the neo-Sanskrit and the neo-Persian elements

(continued on page 15)

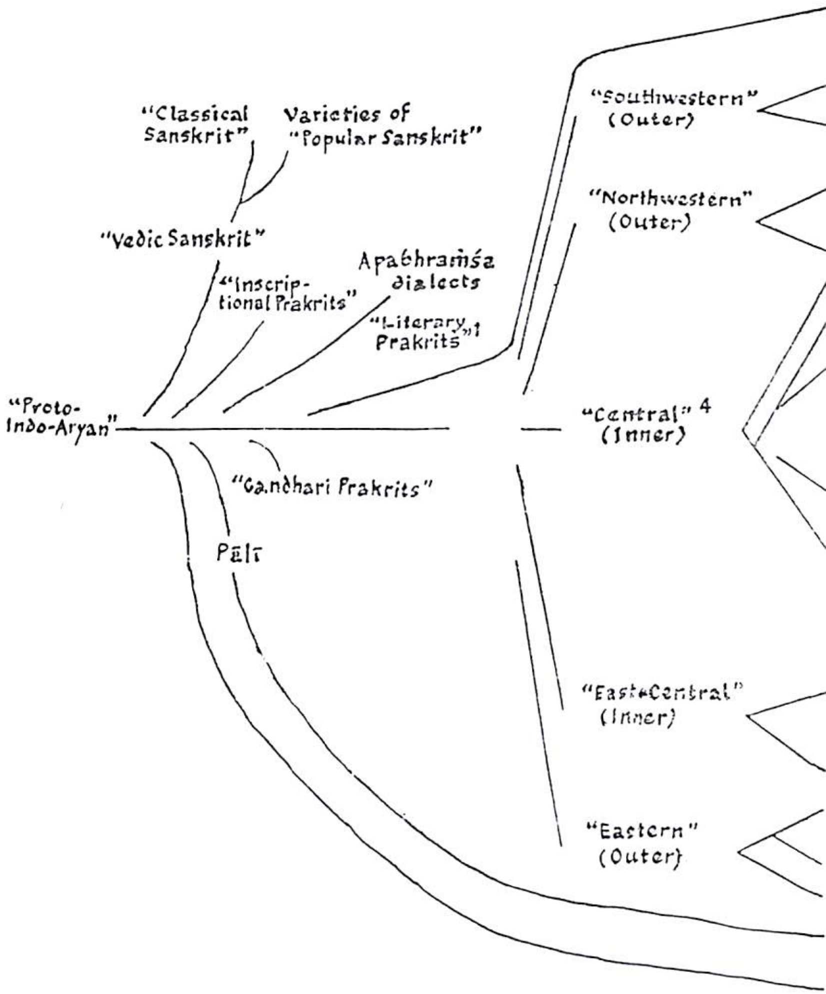


Figure 3. The Place of Hindi-Urdu among

The diagram has no time scale. Artificial labels are enclosed in quotation marks.

The affiliations marked by lines represent the tentative general consensus among scholars today, who recognize serious gaps in our knowledge (some of which show in the diagram) in this matter (especially regarding the time-depths) and who recognize that any such "diachronic" picture has to be corrected by a "diatopic" picture such as the one given in Fig. 1.

Romānī [2] ("Gypsy Dialects")

Komkaṇī

Marāṭhī (including "Marāṭhī Bhīlī")

Sindhī (including Kachchhi [3])

Lahndā (W. Panjab; including Multānī)
Panjābī (E. Panjab; including Ḍogrī)
Gujarātī (including "Gujarātī Bhīlī")

"Rājasthānī"— "West R."(Mārvāṛī, Mevāṛī),
— "East R."(Ḍhūṁḍhāṛī, Haṛautī)
— "North-East R."(Ahīrvāṭī, Mevātī)
— "South R."(Mālvī, Nīmāṛī, Bhīlī proper)

ʼahāṛī [5] — "West P.", "Central P."
— Nepālī ("East P.")

Madhyadeshī" [6] — Bundelī, Bāṅgarū,
Khaṛī Bolī (including "Regional Hindustānī",
"Regional Dakhanī", and "Hindī- Urdū"),
Braj, Kanaujī

Koshalī" [6] — Avadhī (including Baisvāṛī),
Baghelī, Chhattisgaṛhī

Bihārī "— Bhojpurī, Magahī, Maithilī

ʼengali (Bāṅglā)

ʼssamese (Asamīyā)

ṛiyā

inhalese [7] (Siṁhaḷabasa or Heḷubasa)

ʼardic"— Kāshmīrī, Shiṇā, and others

he Indo-Aryan languages.

"Popular Sanskrit" is a cover term for the Sanskrit of the epics (*Rāmāyaṇa* and *Mahābhārata*) and the Purāṇas, of Buddhist and Jain religious writings, and of Tantric and technical writings; this is essentially Sanskrit with Middle Indo-Aryan interference (cf. Late and New Latin). "Inscriptional Prakrits" of course include the various Ashokan dialects. "Literary Prakrits" include the Prakrits of literary prose and verse, of Jain religious writings, of passages embedded in Classical Sanskrit drama; and the Prakrits

described by the grammarians. "Gāndhārī Prakrits" is a cover term for the Prakrits of the various writings in Kharoṣhṭhī script discovered in Afghanistan and Central Asia—notably the Gandhari *Dhammapada* and the Niya documents. "Literary Prakrits" betray the greatest interference from Classical Sanskrit. Apabhraṁśa dialects include the varieties of Apabhraṁśa seen in Sanskrit drama, in literary prose and verse, and in Jain religious writings; the varieties described by the grammarians; as well as the Avahaṭṭha dialect of *Kīrtilatā* and *Raṇamallachanda* sometimes distinguished from Apabhraṁśa proper.

Notes to Figure 3 :

(1) The "regionalisms" of the "Literary Prakrits" and Apabhraṁśa are now seen to be what they really are, namely, interferences from the regional casual speech-forms (comparable to 3A in Fig. 2) and not dialects identifiable in any way with Proto-Southern, Proto-Northwestern, Proto-Central, etc. as was once thought to be the case. (2) Romani has affinities with the "outer Northwestern" group. (3) Kachchhi speakers are now adherent speakers of Gujarati. (4) In the "inner Central" group most probably belongs Afghoni, spoken near Stalinabad in Soviet Central Asia, its exact place being yet undetermined. [This dialect, called ləfz əfɣoni or zəboni əfɣoni by its speakers, was discovered in 1954 by I. M. ORANSKI. For a second-hand account, based on ORANSKI's papers in Russian, see J.-M. PERRIN, L'Afghon, dialecte indo-aryen parlé au Turkestan, *Bulletin de l'E'cole française d' extrême orient* 52:1 (1964) 173-81.] (5) The position of the Pahari sub-group within the "inner Central" group should perhaps be further looked into. Nepali is also called Parbatiyā, Gorkhālī, or Khas-kurā and is of course to be distinguished from the Tibeto-Burman language Newārī, which is spoken in the same area and which is also called Nepali by some. (6) It has been suggested with some reason that in order to reduce the semantic load on the term Hindi the groups hitherto called "W. Hindi" and "E. Hindi" may be renamed. For "E. Hindi" S. K. CHATTERJI's "Koshalī" may be adopted. For "W. Hindi" I venture to endorse Dr. Gyan Chand JAIN's suggestion—"Madhyadeshī". Baṅgaru and Khari Boli have been treated by some as a sub-group within "Madhyadeshi" under the name "Kauravī". Kanauji is regarded by some as a variety of Braj. (7) Sinhalese was once thought to have affiliations with the "outer Southern" group. Now it has been shown, however, that the three-gender system of Marathi is descriptively quite different from and historically quite unrelated to the gender system of Sinhalese.

have mostly come via Persian. Borrowings from Portuguese have perhaps come sometimes via other Indo-Aryan koines. Borrowings traceable to all other languages outside South Asia (with occasional exceptions from French) come via English. Borrowings traceable to all other languages within South Asia are remote and belong to stages antedating "Madhyadeshi".

The more or less undisputed sway of Hirdu in "Upper India" (allowing for the more or less marginal cases from groups 2 and 3 noted towards the end of § 0.1) is of a comparatively recent date (Fig. 4). Hindi literature is barely 80 years old. Urdu is older only by a generation or so. This fact is, however, often lost sight of due to GRIERSON's use of the term "Hindi (Western and Eastern)" (which thus includes Braj, Avadhi, etc. by the side of Khari Boli), due to the institutional inseparability of "adherent speakers" of Hirdu from its "native speakers", and, more importantly, due to the parcelling out of older literary traditions as the respective heritages for Hindi and Urdu (Fig. 4). Old Maithilī, Old Koshalī, Old Baghelī, Old Avadhī (including Old Baisvāṛī), Old Braj, and finally Old Khaṛī Bolī (the last entering into the Sadhukkaṛī of Kabīr and others and appearing in certain passages from Bhūṣaṇ) are claimed by Hindi. So a history of "Hindi" literature will embrace all of the foregoing. Correspondingly a history of "Urdu" literature will cover Hindavī or Rekhtā (some prefer to call it Old Urdu) and Old Literary Dakhani. It may be noted that Amīr Khusrau, some of the early printed books of Lallū Lāl and others, and Premchand are claimed both by Hindi and Urdu; that Vidyāpati (who wrote in Old Maithili and in Apabhraṁśa) is claimed both by Hindi and Bengali;

(Continued from page 11)
are distinct from classical Sanskrit in its heyday and from classical Persian and Modern Persian. The neo-Arabic elements, again, are unrelated to classical Arabic and colloquial Arabic dialects as these are used today.

that Nānak (who wrote in Old Braj, Old Khari Boli, and Old Panjābī) is claimed both by Hindi and Panjabi; and that Mirā̃ (who wrote in Old Braj with Middle Rajasthani interference) is claimed both by Hindi, the proponents of Modern Literary Rajasthani, and Gujarati. Old (West) Rājasthānī and Middle Rājasthānī constitute the heritage of the proponents of Modern Literary Rajasthani. Old West Rajasthani (under the name Old Gujarāti) and Middle Gujarātī (1450–1800, Narsī Mahetā onwards) constitute the heritage of Modern Gujarati. Old West Rajasthani and Middle Rajasthani as used by the bardic poets are sometimes called Ḍiṅgaḷ and opposed to Piṅgal (Old Braj as used in verse by Rajasthani speakers).

0. 3. Methodological preliminaries.

If any single thing is clear about Hirdu, it is that nothing is clear—or, less pessimistically, the situation of Hindi-Urdu is enormously complex in all the dimensions—linguistic, socio-cultural, and literary; diachronic (historical) as well as diatopic (geographical). From now on we can concentrate on contemporary standard Hindi-Urdu (Fig. 2, nos. 1, 2, 3), which is itself a continuum of styles polarized between Hindi (Fig. 2, no. 1A) and Urdu (Fig. 2, no. 1B). Regional colour (Fig. 1, no. 3A) from the "lower languages" (§ 0.1, groups 1 and 2, but not groups 3, 4, and 5) will be "taken care of" in the description but not precisely located geographically—partly because this is by no means a definitive study and other people are working on it and partly because each idiolect—whether "native" or "adherent"—shows its own brew of regionalism(s) (speakers of the standard Hindi-Urdu come from a highly mobile section of the population) and well-defined uniformities are not detectable. One accidental limitation must be noted—I cannot guarantee that all Pakistani regionalisms of

Urdu will be covered in this study; I have had little or no access to speakers of Urdu in Pakistan.[10]

An adequate description of contemporary standard Hindi-Urdu (from now on called Hirdu) should enable us, then, not only to produce all and only phonologically, grammatically, and semantically well-formed utterances but also to assign adequate stylistic indices to such formulae. So, out of such an overall description, one can pull out certain traits and make up any given idiolect. If one is thinking of adopting the "generating protocol" (a formalized set of instructions or rules for generating) as a mode of description, one can say that the rules will generate well-formed utterances, but the meta-rules manipulating the stylistic indices will generate homogeneous speech-varieties (or "microlects", as I prefer to call them). In the present study we shall use a rough and ready apparatus that will be "delicate" enough (to adapt HALLIDAY's term to a slightly novel use) for most purposes. We shall use the adverbial tags **H**, **U**, and **HU**. The last one will be left understood when the context does not call for an explicit symbolization. We shall also use the following symbols: ⊙ for nothing, $\sim$ for a free variant, $\approx$ for a familiar or less elegant variant (including hypercorrect forms—described as such), ∞ for an allegro variant, and $\simeq$ for an extra-elegant form. The microlect selecting all and only **H** and **HU** alternants will probably be an extreme version of Hindi usable only in a formalized highbrow or middle-

10. I have heard it said that since the Partition of 1947, Urdu in India has come closer to Hindi and Urdu in Pakistan has diverged even further from Hindi so that even Indian speakers of Urdu have some difficulty with it. I have also heard that some of the Panjabi regionalisms have now become the prestige forms in Pakistani Urdu. Both of these observations about Pakistani Urdu are said to be particularly true of the language of Radio Pakistan and of Pakistan Government Publications.

Pak. ✓

brow style. Similarly the microlect selecting all and only **HU** and **U** alternants will be an extreme version of Urdu usable only in a formalized highbrow or middlebrow style. Finally there will be the microlect selecting all alternants other than **H** $\simeq$ and **U** $\simeq$ alternants—an extremely balanced version of Hindustani usable probably only in a formalized or casual middlebrow style. Any text showing a coexistence of **H** $\simeq$ and **U** $\simeq$ items betrays the presence of code-switching on the part of the speaker (or writer) of that text. A text exclusively in one of these three rather idealized microlects is probably not a very common occurrence. Regionalisms in this scheme will be distributed between **H** $\approx$, **U** $\approx$, and **HU** $\approx$ taggings. No further refinement will be attempted in this study.[11] It should perhaps be made clear at this point that **H**, **U**, **HU** are in the first instance descriptive devices and have no reference either to linguistic history or to linguistic sociology or geography, though such correlations could be attempted in some cases.[12]

The purpose of the other set of tags used in the present study—Sanskrit, Persian, and English—is chiefly a rapid identification that will enable the Hindi-Urdu-knowing reader to know where to look for more examples of the same kind. Historically a "Sanskrit" word may have come via Bengali, or a "Persian" word may have

11. For example, one might attempt to set up not just **H** and **U**, but $\mathbf{H_1}$, $\mathbf{H_2}$ and $\mathbf{U_1}$, $\mathbf{U_2}$ so that $\mathbf{H_1}$-$\mathbf{U_2}$ (or $\mathbf{H_2}$-$\mathbf{U_1}$) collocations in the same idiolect are expected to be more probable than $\mathbf{H_1}$-$\mathbf{U_1}$ ones but less probable than $\mathbf{H_2}$-$\mathbf{U_2}$ ones. This way $\mathbf{H_1}$, $\mathbf{H_2}$, **HU**, $\mathbf{U_2}$, $\mathbf{U_1}$, form a connected series with $\mathbf{H_1}$ and $\mathbf{U_1}$ as the two extremes.

12. Applying the tags **H**, **HU**, **U** to lexical items, we may perhaps be able to say that Sanskrit loanwords will be typically **H** $\simeq$, **H**, or **HU**, or that **U** $\simeq$ items are more likely to be Persian loanwords traceable to Arabic than those not so traceable. On the socio-geographical plane, one can make similar statements about the speech of, say, Muslims in the "Bihari" area, or Kayastha Hindus in the Braj area.

come from Arabic ultimately, or an "English" word may be traceable to Burmese for all we know. The synchronic unification of these strata is helped by the presence of cognates like dur 'far' in Sanskrit durdəršī 'far-sighted' and Persian durəndeš 'far-sighted'.

The glosses are intended to be rough and ready identifications—either by providing translation equivalents or, in parentheses, explanations. (Thus xɪtab *means* a 'title', but šex *is* '(a title)'.)

0. 4. Recommendations on nomenclature.

The various scattered recommendations on nomenclature may usefully be brought together here : In the linguistic (as well as cultural) area of *South Asia* (don't loosely call it India, as some do), are found various language families. *Indo-European* is represented by the sub-family *Indo-Aryan* (otherwise Indic or Gaudian). Some of the *contemporary* (which has a shallower time-depth than *Modern*, which in South Asia means after Western contact) Indo-Aryan dialects fall within the *Hindi-Urdu adherent area*—they are : dialects of the *Madhyadeshi* (not Western Hindi), *Koshali* (not Eastern Hindi), *Bihari*, Western and Central *Pahari*, *Rajasthani* (in the context of literature Rajasthani refers to Marvari, its Western dialect), and *Lahnda* (not Western Panjabi) groups. The Madhyadeshi group is constituted by *Khari Boli*, Bangaru, Braj (including Kanauji), and Bundeli (Kauravi is not recommended). Contemporary Khari Boli covers, besides *Hindi-Urdu* (*Hirdu* for short), *Regional Hindustani* (not Vernacular Hindustani) and *Regional Dakhani* (different from Old Literary Dakhani). Hindi-Urdu is essentially non-regional, that is, regionally unmarked. It has standard as well as substandard varieties : the substandard forms are lowbrow casual and not ordinarily reduced to writing. The standard forms use the *Devanagari* and *Perso-Arabic scripts* (don't call the

scripts Hindi and Urdu, as some do) : *Hindi* (by definition standard, formalized highbrow and middlebrow that is H-oriented) uses the former, *Urdu* (by definition standard, formalized highbrow and middlebrow that is U-oriented) uses the latter, *Hindustani* (HU-oriented, never highbrow) uses both. Hindustani may even be substandard (low-brow formalized).

Urdu-heritage literature includes Urdu, *Old Urdu* (perhaps more convenient than Hindavi or Rekhta), and *Old (Literary) Dakhani* (literary may be dropped where no contrast with Regional Dakhani is intended) literatures. *Hindi-heritage literature* includes Hindi and all the remaining literatures of the Hindi-Urdu adherent area (currently a large part of Old Braj Literature written by Rajasthani writers and Old and Middle Rajasthani literatures happen to be ignored, however). *Old Gujarati–Rajasthani* may be suggested as a suitable label to replace Old Rajasthani, Old West(ern) Rajasthani, Old Gujarati, Old Marwari.

It will be noticed that only the recommended terms are italicized. The longer, clumsier terms must not be abbreviated whenever precision is called for. Informally of course the words "adherent" and "heritage" may be left understood. Those who do not like the sound of the term "Hirdu" are of course quite welcome to use the fuller form "Hindi-Urdu".

The present recommendations may or may not turn out to be acceptable. The one desideratum that may be widely accepted is that the present chaos be ended.

Chapter 1

WORD PHONOLOGY

1. 0. Introductory.

The phonologic word is the maximal domain of this study and may be defined as the stretch of segmental phonemes with no non-close junctures within and bounded by non-close junctures and/or utterance boundary. The terms initial, medial, final will refer to this unit. No attempt will be made here to establish non-close junctures—# will be used as a generalized cover symbol for a word boundary. (It is hoped that this could be undertaken in a future study). Occasionally /+/ for word juncture is used in some examples consisting of a sequence of phonologic words.

The contrasts within a Hirdu word involve segmental phonemes—consonants and nonsyllabic and syllabic vocoids—and one coarticulation phoneme of nasalization.

The study opens with a sketch (§ 1.1) of the core system—embodying the contrasts which seem to be firmly established (in most cases with a good functional yield) and to enjoy a fair degree of uniformity over the whole Hirdu field.

Then follows a rather detailed treatment (§ 1.2) of the problem spots. The grouping is for ease of presentation—loanword consonants, nasality, retroflex flaps, glottal friction, marginal vowels, semivowels and diphthongs, clusters and cluster-supporting vowels, and atonic short vowels. There is no logical consistency or ordering in this; but it is hoped that this will make it easy for the reader to follow the description.

After gathering a few miscellaneous correspondences, (§ 1.2.9), I have proceeded to present a consolidated

picture (§. 1.3) and draw some implications (§ 1.4) of a practical or theoretical interest.

1. 1. A Sketch of the core system of word phonology.

1. 1. 0. Inventory of phonemes.

To start with, we may recognize a core system of 28 consonants, 3 nonsyllabic vocoids, 8 syllabic vocoids, and 1 coarticulation. (See Table 1.) It is of course to be understood that this sketch is by no means intended to cover the whole territory : it only maps those areas about which we can be reasonably certain and thus establishes a base of operations for § 1.2.

TABLE 1

Consonants (C) :					
Stops	p	t	ṭ	č	k
	p^h	t^h	$ṭ^h$	$č^h$	k^h
	b	d	ḍ	ǯ	g
	b^h	d^h	$ḍ^h$	$ǯ^h$	g^h
Nasals	m	n			
	m^h	n^h			
Sulcate		s			
Laterals		l			
		l^h			
Tremulant		r			
Nonsyllabic vocoids (W) :	h				
		y			v
Syllabic vocoids (=vowels, V) :		i			u
			ɪ	ʊ	
		e			o
			ə		
			a		

Coarticulation : ^{n} (nasalization)

1. 1. 1. Contrasts illustrated.

pəla 'was reared', təla 'sole of the foot; deep-fried', ṭəla 'was displaced or avoided', čəla 'walked, set out', kəla 'art'; p^{h}əla 'bore fruit', t^{h}əka 'was tired', ṭ^{h}əga 'cheated', č^{h}əna 'was strained; have (it) strained (imper.)', k^{h}əra 'genuine'; bəla 'calamity', dəla 'crushed', ḍəla 'was put into', ǯəla 'burned (intrans.); burn (transitive, imperative)', gəla 'throat'; b^{h}əla 'good-hearted', d^{h}əra 'caught', ḍ^{h}əla 'was spilled', ǯ^{h}əla 'swayed', g^{h}əṭa 'heavy clouds in a group'; məla 'rubbed', nəla 'long tubelike bone'; səta 'annoy (imperative)', ləga 'touched', rəṭa 'repeated orally'; hərа 'green', yəha^{n} 'here', vəha^{n} 'there'; əda 'gesture, manner'.

čači 'father's brother's wife', atši 'fireproof'.

kumar 'boy; male given name', sunar 'goldsmith'; kum^{h}ar 'potter', ǯun^{h}ai 'moonlight'; gulal 'red powder used in Holi', kul^{h}a 'half of rump, buttock'.

mil 'mile', mul 'root', mɪl 'meet (imperative)', pul 'bridge', mel 'process of meeting, harmony', mol 'price', məl 'rub (imperative)', mal 'commodity'.

əč^{h}ut 'untouchable', č^{h}ut 'pollution by contact'.

kəha 'said', kəha^{n} 'where?'; sas 'mother-in-law', sa^{n}s 'breath'; čaṭa 'licked', ča^{n}ṭa 'slap, blow'.

1. 1. 2. Phonetic correlates.

The basis of articulation: **H** fronted articulation on the whole, clearer quality; **U** retracted articulation on the whole, darker quality.

The basic values: s r are alveolar; the rest in that column (in Table 1) are dental. The retroflex consonants are: **H** fronted, post-alveolar, **U** retracted, cacuminal. The affricates have a long-groove lamino-dorso-alveolar-pre-palatal [š]-type release, **H** clear to neutral in quality, **U** neutral to dark in quality. The occlusion is

H lamino-alveolar, U apico-lamino-alveolar. r is weakly trilled unless geminated; U≃ [ɹ] in —V, V—#. The consonants in the first two rows and the sulcate are voiceless and fortis; the rest are voiced and lenis. Superior h indicates aspiration, voicelss or voiced according to the consonant. The vocoids are voiced. i e u o are near-cardinal and tense; i e u o and ɪ ʊ are : H more peripheral, U e o slightly lower; i u ɪ ʊ more neutral. ɪ ʊ ə a are lax. ə is : H low-mid central slightly retracted; U low-mid central slightly higher and fronted. a is low central. v u ʊ o are rounded; y i ɪ e are unrounded with spread lips; ə a are unrounded. The peripheral vowels i u e o a (symbol $\bar{V}$) are progressively longer than the neutral vowels ɪ ʊ ə (symbol $\check{V}$). Superior n indicates full nasalization of the preceding syllabic vocoid together with any nonsyllabic vocoids or lateral that may be tautosyllabically adjacent to it. In phonemic writing it will be placed after V or VW as the case may be (W being any nonsyllabic vocoid).

Positional variants : n l are alveolar when adjacent to retroflex consonants. r is more fronted after dental consonants. All voiced stops are prenasalized after V^n or VW^n—that is, they have a short nasal consonant as an on-glide. Intervocalic m m^h are : HU ∞ extra-lenis with imperfect oral occlusion. r is somewhat lateralized before l and fricativized after voiceless consonants. The aspiration of all voiced aspirates is extra-lenis syllable-finally in HU∞. The vocoid h has a voiceless on-glide initially. In V–C, h has a vocalic off-glide echoing the preceding syllabic vocoid. Finally h is : H with a vocalic off-glide, U with a voiceless off-glide. A vowel accompanied by /^{n}/ is slightly higher. A vowel in a free (open) syllable when followed by another vocoid other than h is higher than in a checked (closed) syllable. A vowel adjacent to a tautosyllabic retroflex consonant is slightly retroflexed—also, ɪ ʊ ə are lower and a is more retracted in this

environment. A vowel is slightly breathy before a tautosyllabic voiced h—in **HU** ∞ the vowel is fully breathy and the h is not a separate segment. The problem of ḍ ḍʰ will be taken up later—for the present we may simply say : initially, in gemination, and before a final y or v we *always* find [ḍ], [ḍʰ] (stops, the geminates being [ḍḍ] or [ḍḍʰ]) and elsewhere we *generally* find [ṛ], [ṛʰ] (flaps).

Syllable division : For many purposes (cf. the use of the term "tautosyllabic" above and the terms "free" and "checked syllables" above and immediately below) it is useful to set up phonologic syllables within a word. Needless to say, this division is entirely predictable ; otherwise there would be junctural contrasts within a word. The procedure for this is as follows : Having written the word phonemically, grade the segment to segment transitions in the following descending order of closeness—

1 CV, hV, yV, vV
2 VC, Vh, Vy, Vv
3 Cy, hy, vy
4 Cr, Cl, Cv (except l̃l, lr, lv, rl, rr, rv)
5 NC* (where C* is any stop or sulcate and N is a nasal homorganic with C*)
6 all transitions other than 1–5, 7, 8
7 all geminate nonsyllabics
8 Ch

All nonsyllabics before the first vowel or after the last vowel go with that vowel. Between two adjacent vowels there is always a cut. In a sequence of nonsyllabic(s) flanked by two vowels, cut at the openest transition. If there are more than one such, cut at the first out of them. Examples of phonologic syllable division :[13]

13. For more examples see the next subsection on the Placement of tonic and subtonic.

svəst^{h} 'healthy'
b^{h}ai 'brother'
so^{2}.t^{1}a 'sleeps'
so^{2}.y^{1}a 'slept'
ə^{2}s^{6}. v^{1}əst^{h} 'unwell'
ʊ^{2}t^{4}. r^{1}a 'descended'
mə^{2}n^{5}. t^{4}r^{1}i 'minister, secretary'
sə^{2}n^{5}s^{6}. k^{4}r^{1}ɪt 'Sanskrit'
mɪ^{2}t^{4}r^{6}. t^{1}a 'friendship'
pra. tə^{2}h^{6}k^{7}. k^{1}al 'morning'
lə^{2}k^{6}. ṣ^{6}m^{1}i 'Lakshmi, the Hindu goddess of wealth'

Placement of tonic and subtonic : Hirdu has accent entirely as a part of intonation. A word may not have accent in a many-word contour-span. A word has always an accent in a single-word contour-span. The syllable which bears the accent on a word may be called the tonic. Even when the word bears no accent the tonic is slightly more prominent than the other syllables. The subtonic, if any, is the next most prominent. The tonic and the subtonic will be phonetically marked by a preceding '' and ' respectively. The procedure for predicting the tonic (and the subtonic, if any) in a word is as follows : Having marked the vowels as short (ɪ ʊ ə) and long (i e u o a) and having marked the phonologic syllable cuts (as described above), grade these syllables in the following ascending order of heaviness—

Light : a syllable ending in a short vowel;

Medium : a syllable ending in a long vowel or having a short vowel checked by a single nonsyllabic;

Heavy : any other syllable.

(Please note that /^{n}/ is being ignored here.) Find the maximally heavy syllable or syllables. If there is only one maximum, it is tonic; the remaining preceding and

following syllables, if any, are alternately atonic and subtonic, those adjacent to the tonic being atonic. If there are more than one maximum, the last but one out of the maxima is the tonic; the heavy syllable(s) preceding the tonic, if there are any, are all subtonic; the remaining syllables are alternately atonic and subtonic, those adjacent to the tonic or to a heavy syllable, however, being always atonic.[14] Phonetically the prominence of a tonic or subtonic syllable consists of extra-length to the coda nonsyllabics or (in case there is no coda) to the nucleus. Examples of the placement of the tonic and subtonic :

"nə 'not' "L
"mən 'mind' "M
"gal 'cheek' "H

kɪ"d^{n}ər 'which way ?' L"M
ǯə"nab 'sir' L"H
"sav^{n}.la 'tawny' "HM
əs"bab 'goods' M"H

sʊ"ča.rʊ 'attractive' L"ML
"šox.ǯə'ba.ni 'talkative' "HL'MM
'ɪn.qɪ"lab 'revolution' 'ML"H
'mʊ.səl"man 'Muslim' 'LM"H

sə"mɪ.tɪ 'committee' L"LL
ro"za.na 'daily' M"MM
"səns.kar 'impression' "HH
"roz.gar 'employment' "HH

---

14. Considering the varied failure of previous attempts at the formulation of these rules, the present attempt may be excused if it looks somewhat complicated. The basic breakthrough here was Dr. JAIN's (see Preface) brilliant suggestion that we need three degrees of "heaviness" correlated with the "length" of the vowel and of its consonantal coda.

"re.ḍɪ'o 'radio' "ML'M
rʊ"ka.ya 'stopped (tr.)' L"MM
ka"ri.gə'ri 'craftsmanship' M"ML'M

'as"man.ǯah, U ≃ "as.maⁿ'ǯah 'highly placed' 'H "HH, but "HM'H
pər'vər.də"gar 'God as protector' M'ML"H
"rez.ga'ri 'small change' "HM'M

It goes without saying that the word here is the phonologic word bounded by plus junctures. Examples of word sequences :

kər"va.uⁿ + "ga 'I shall have (it) done'
"čʰo.ṭa + "sa 'of the small type'
"bʰo.la + "pən 'credulity'
"xʊš + "qɪs.mət 'good fortune'
"səb + "hi 'all (emphatic)'.

In some of these cases + is replaced by a close juncture in allegro style and the placement of the tonic and the subtonic adjusted accordingly. Any further grading as between two neighbouring tonics is beyond the scope of word phonology.

1. 1. 3. Highlights of the distributiodal statement.

A word may begin or end in a vowel. Final vowels, however, are generally the peripheral or long ones. Vowels may follow each other. All nonsyllabics except m^h n^h l^h (and of course ṛ $ṛ^h$) can occur initially. All nonsyllabics can occur intervocalically. All nonsyllabics except m^h n^h l^h can occur finally. All initial and final sequences of nonsyllabics that occur are either **H** or **U** or both, but never **HU**. Intervocalic sequences of nonsyllabics occur with greater freedom and out of the ones that occur some are **H**, some **U**, and some **HU**. V^n is never followed by a nasal consonant in the same syllable.

1. 2. A Detailed survey of problem areas.

1. 2. 0. Introductory.

With the foregoing sketch for a background, we shall now proceed to deal in some detail with cases of marginal contrasts and/or complex cases of **H**, **HU**, and **U** correspondences. In doing so, we shall naturally adopt the solution which among other things modifies the sketch of the core system (in § 1.1) the least. The relevant allophonic and distributional statements will be sketched in at appropriate points.

No attempt will be made to work out the morphophonemic statement—such as will enable one to rewrite phonemically a morphophonemically written utterance. Enough hints, however, have been provided if anyone is so inclined. One general hint may also be thrown out here. It is clearly essential to make a distinction between the ascription of **H**, **HU**, **U** tags to morphemes and vocabulary items in a lexicon and to the ascription of these to a piece of phonemic writing (together with its phonetic rendering). (See, for example, Table 2.)

TABLE 2

The Lexical items :		Their phonemic-phonetic rendering :		
H	{bhasma}	**H**	/bhəsm/	["b^{h}əsmə̆];
		U	/b^{h}əsəm/	[b^{h}ə"səm]
HU	{rupee}	**H**	/rʊpya/	["rʊpĭ.ya];
		U	/rupɪa/	[rʊ.pɪ"^{y}a]
U	{health}	**H**	/sehət/	["se.hət];
		U	/sɪhhət/	["sĕẖ.hət]

Some of the correspondences between **H**-phonemic and **U**-phonemic spellings will be too sporadic to permit useful systematization—as with the **HU** item {ant} : **H** /či^{n}ṭi/, **U** /čyu^{n}ṭi/. But many of them will call for a systematic statement. Such a statement will be very

closely meshed in with the more conventional kind of morphophonemics concerned with positional and free variants (the latter inclusive of style and tempo variants). For example, even **U** has /b^{h}əsm-/and not /b^{h}əsəm/ in /b^{h}əsmo^n/. So, for each lexical item we can provide one or rarely more morphophonemic spellings. A morphophonemic spelling will, in most cases, be **HU**; some of the inevitable exceptions will be cases of sporadic correspondence. The conversion rules will not permit **U**-phonemic spellings to be derived from **H**-morphophonemic spellings and vice versa. Otherwise there will be no restrictions. Occasionally there will be correlations between the original tag of the lexical item and the choice of the phonemic rendering. For example, the form **U** /b^{h}əsəm/ will coexist with **U**/qɪsm/ (from **U** {type}) and the form **H** /b^{h}əsm/ will coexist with **H** /kɪsəm/—simply because {bhasma} is **H** and {qism—type} is **U**. Perhaps it will be best to work this kind of thing into the morphophonemic spelling itself (in this case, by spelling the last portion differently for {bhasma} and {qism-type}), so that the conversion rules can be applied blindly without worrying about the original lexical tag or, for that matter, about the morphophonemic tag. There will be no change of tag in converting a phonemic spelling into a phonetic one.

The following general stipulation is to be understood: "**H** + **HU**" style is more apt to use familiar variants for Persian loanwords than "**HU** + **U**" style would, and "**HU** + **U**" style is more apt to use familiar variants for Sanskrit loanwords then "**H** + **HU**" style would be.

Since the material that follows has been arranged by problems, the spellings of the examples for a given problem may sometimes anticipate solutions to problems presented later. The reader should be rather patient in such cases. The correspondences have been numbered

consecutively for ease of reference. (All references where the number stands alone are to these sets from 1 to 117 in § 1.2.)

1.2.1. The Problem of the marginal loanword consonants.

These may be conveniently grouped into Sanskrit, Persian, and English. The correspondence-sets under each are arranged roughly in an order of decreasing effectiveness of foreign interference and increasing assimilation to the native patterns.

Sanskrit items call for the use of v(90), y(89); š (1), ṇ (21); ḍ (25) intervocalically and elsewhere; ṣ (2), [ḫ] (the visarga; 28, 29).

(1) **H, HU** š ≈ s, **U** š, as in : šantɪ 'peace', sənšəy 'doubt', prəveš 'entry, entrance", prəšn 'question', pərdeši 'foreign, foreigner'.

/š/ [š] is homorganic with the affricates /č/ etc.

(2) **H** š ≃ ṣ≈s, **HU** š ≈ s, **U** š, as in: ṣəṭkoṇ 'hexagon', b^haṣa 'language', səntoṣ 'pleasure', səntuṣṭ 'pleased'.

/ṣ/ [ṣ] is not a true retroflex, but more like [š] with the apex retracted and raised and with darker quality.

Persian items call for the use of v (90, 91), y (89); š (3, 4); f (5), z (6); x (7); ɣ (8), q (9); ž (10), ʾ (11). Note that the voiceless pharyngal fricative of Arabic is always /h/ in Urdu.

(3) **H, HU** š ~ s, **U** š, in a few items, as: šamɪl 'included, acceded', šəmšer 'sword' (the first š), šəlvar 'kind of loose trousers'

(4) As with (1), in most other items, as: šah 'emperor', xuši 'pleasure', hoš 'consciousness', šəmšer 'sword' (the second š), gošt 'meat'.

(5) **H** p^h, **HU, U** f, as in fɪkr 'worry', kafi 'enough', saf 'clean', muft 'free of charge', fəran 'at once', əfyun 'opium' (also əfim 'opium' of doubtful origin).

/f/ [f ~ ɸ] adjacent to tautosyllabic back rounded vowels, [f] elsewhere

(6) **H** ǯ, **HU**, **U** z, as in : zor ' force ', zɪndə ' alive ', čiz ' thing ', mɪzaǯ ' health, temperament ', roz ' day ', vəzn ' weight '.

/z/ **HU** [^{d}z], **U** [z], the voiced counterpart of /s/.

(7) **H** kʰ, **HU** x ~ kʰ, **U** x, as in : xʊd 'self', xɪtab 'title', naxun 'fingernail', šox 'fickle', səxt 'hard', xvab 'dream'.

/x/ [x] rather retracted and with uvular scrape.

(8) **H, HU** g, **U** ɣ, as in : ɣəm ' sorrow ', mʊɣəl ' inhabitant of Turkestan in Central Asia ', baɣ ' garden ', bəlɣəm 'mucus, phlegm as a humour in Yunani medicine'.

/ɣ/ [ɣ] even more retracted than/x/.

(9) **H, HU** k, **U** q ~ x before a voiceless consonant, q ≈ x elsewhere, as in : qatɪl 'murderer', baqi 'remainder' təlaq ' divorce ', əql ' intelligence ', vəqt ' time ', vəqf ' property dedicated to God '.

/q/ [k̩] a voiceless post-velar stop.

(10) **H** ǯ, **HU** z, **U** z ≃ ž, found in a small number of items, as : mɪžɔ ' eyelid, eyelash ', mɪžgan ' ditto, plural ', žaž ' (a thorny shrub); idle or obscene talk ', žalɔ ' hail, dew, hoarfrost ', aždʰat ' bell metal ', pɔžmʊrdɔ ' withered '.

/ž/ [ž] the voiced counterpart of /š/.

(11) **H, HU** ⊙, **U** ⊙ ≃ ʔ, corresponding to 'ain in Persian orthography and especially liable to be dropped syllable-finally, as in ʔɔrət ' woman ', bəʔid ' far ', ɪnʔam 'reward', məsʔud ' beneficent ', bʊrqaʔ 'woman's robe fitted with a veil ', taʔnə ' taunt ', saʔd ' good, holy ', fɪʔl ' deed, verb '. Rarely ʔ in C—V corresponds to hamzā in Persian orthography, as in : ǯʊrʔət ' daring ', qʊrʔan the Qur'ān '.

/ʔ/ [ʔ] fortis glottal stop in #—V, lenis glottal stop or constrictive elsewhere. The use of /ʔ/ is now very

definitely a mark of affectation rather than elegance; its use is practically limited to the speech of some maulavis.

Vowels adjacent to a tautosyllabic ɣ, q, or ʔ are accompanied by a faucal constriction, especially U ≃. (See also 30, 44, 45, 116, 117 below.) Note the placement of the tonic in "bʊr.qa ≃ bʊr"qaʔ above and syllable division in ɪ"nam ≃ ɪn"ʔam and ʒʊ"rət ≃ "ʒʊr.ʔət above.

English items call for the use of v (90), y (89); š (12); f (13), z (14); ḍ (29). Note that English ž is always ʒ in Hirdu (as in : ruʒ 'rouge', ṭelɪvɪʒən 'telivision').

(12) As with (1), for example in: šo 'display, pretence', sṭešən 'railway station', kɛš 'cash'.

(13) H $p^{ɦ}$; HU, U f ~ p^h, as in : fis ≃ fi 'fee', kafi 'coffee', sṭaf 'staff, body of workers'.

(14) As with (6), for example in : zon 'zone', čiz 'cheese', kəmyunɪzm 'communism'.

1.2.2. The Problem of nasal consonants and nasalization.

The contrast between V^nC^* and VN^*C^* (where C^* is any stop and N^* is a nasal, [m n ṇ ŋ], homorganic with that stop) is unmistakable phonetically in spite of partial similarities.

/bɪⁿdʰa/ 'made a slit' [bɪ̃ⁿ"dʰa].
/hɪndi/ 'Hindi' ["hɪ̨n.di].
/daⁿt/ 'tooth' ["dã̂t].
/šant/ 'quiet, peace' ["šą̣nt]

(Note the prenasalization of d^h, the placement of the tonic, the raising of vowels with /ⁿ/, and the partial nasalization shown by a subscript hook of vowels followed by the nasal consonant.) But distributionally it is an unstable contrast. VN^*C^* is chiefly illustrated by Sanskrit, Persian, and English loanwords.

gʊmpʰa ~ gʊpha 'cave, grotto', kʊmbʰ 'earthen waterpot, Aquarius', ʒənta 'people (in a political sense)', mənd

'quiescent, slow', nɪnda 'disparagement', sɔndəry 'beauty', akʊnčən 'contraction', paṇḍity 'erudition', šəŋk^h 'conch'.

gʊmbəz 'dome', čənd 'a few', zɪndə 'alive', xəzanči 'treasurer', čʊŋgi 'duty, tax', rəŋgin 'colourful, coloured'.

ḍəmbel 'dumbbell', bəṇḍəl 'bundle', sɪmeṇṭ 'building cement', bɛŋk 'bank (commercial)'.

gand^{h}i '(surname)', manta 'accepts', sʊnta 'listens', čəndel '(name of Rajput clan)', k^{h}ɔnča 'platter used by pedlars', b^{h}anǰa 'sister's son', sənǰona 'to put together', g^{h}əmənḍ 'conceit', əṇṭsəṇṭ 'this or that, unimportant', beḍ^{h}əŋga 'clumsy', gʊŋga 'dumb'.

Note that $\bar{V}N^*C^*$ is not so common. V^nC^* is not so limited in occurrence as VN^*C^* is. $\check{V}^nC^*$ is less frequent than $\bar{V}^nC^*$.

sə^{n}b^{h}alna 'to take care of', sa^{n}p 'snake', ta^{n}ba 'copper', ǰɔ^{n}čta 'becomes, suits', sɪ^{n}čai 'irrigation', pu^{n}ǰi 'capital', pu^{n}č^h 'tail', ča^{n}di 'silver', bə^{n}ṭa 'was divided; cause to divide (imperative)', čə^{n}ṭ 'cunning', g^{h}u^{n}ṭ 'mouthful (of liquid)', ba^{n}ṭ 'divide, distribute', ʊ^{n}gli 'finger', rə^{n}gana 'to colour (tr.)', sɪ^{n}g^{h}aṛa 'water chestnut, Trapa bispinosa, Roxb.', ḍ^{h}o^{n}gi 'hypocritical', kəru^n+ga ∞ kəru^{n}ga 'I shall do'.

Items in this second group tend to have familiar variants with VN^*C^*.

(15) $V^nC^* \approx VN^*C^*$. Increasingly liable to this alternation are $\bar{V}^nP$, $\check{V}^nP$, $\bar{V}^nB$, $\check{V}^nB$ where P is a voiceless stop and B a voiced stop. With some items and in some idiolects these familiar variants remain the only ones in use, as in gəṇṭ^{h}a 'onion bulb', čand 'moon', ɔmbɪa 'raw mango' rather than gə^{n}ṭ^{h}a, etc.[15]

15. Historically this process is undoubtedly helped by cognates between this group and the various loanword groups with NN^*C^* : re^{n}gana 'to colour', Sanskrit rəŋgɪt 'coloured', Persian reŋgin 'colourful'; bə^{n}d^{h}na 'to be

The net upshot is the incipient polarization between $\breve{V}N^*C^*$ and $\bar{V}^nC^*$.

(16) $V^nP \approx VP$, as in : ǰə^{n}čta, sɪ^{n}čai, bə^{n}ṭa above.

(17) $\breve{V}^nb^h \infty \breve{V}m^h$, as in : sə^{n}b^{h}alna above.

The contrast between VN^*C^* and $VN^{**}C^*$ (where N^{**} is a nasal *not* homorganic with C^*) is possible but infrequent except where N^{**} is m. This contrast as well as cases of VNC^{**} (where N is any nasal and C^{**} is any nonsyllabic other than a stop) call for the setting up ṇ and ŋ as phonemes. (Cases of $N^{**}C^*$ and NC^{**} involving ṇ as the first member are all Sanskrit loanwords, as are also some casaes of ŋC^{**}. See also 21 below.)

$VN^{**}C^*$: səmta 'equality', ǰəmta 'freezes', sɪmṭa 'shrank', səmǰ^{h}a 'understood'; ənbən 'disharmoy in personal relation', sənki 'whim; whimsical', kanungo '(name of a revenue official)', čɪngari ~ čɪŋgari 'spark'; gʊṇkari 'effective (remedy)'.

VNC^{**} : səmman 'respect', ǰəmna 'to freeze', rəmṇi 'beautiful woman', ɪmli 'tamarind', čəmṛa 'leather', somvar 'Monday'; ǰənma 'was born', sʊnna 'to hear', ɪnsan 'human being', mənzɪl 'destination', vanro^n 'monkeys (oblique pl.)', samany 'common'; mriṇməy 'made of earth', gəṇna 'counting', əkṣʊṇṇ 'not beaten (track)', ərəṇy 'forest', əṇvəstr 'atomic weapon'; vaŋməy 'literature', dɪŋnad 'sound reverberating in all directions', səŋhar 'destruction'.

(18) V^ng [$\tilde{V}$ŋ] before a nasal, V^ng [$\tilde{V}$ŋg ∞ $\tilde{V}$ŋ] before #, v, or ṛ, V^nk [$\tilde{V}k \infty \tilde{V}$ŋ] before a voiceless consonant, V^ng [$\tilde{V}$ŋg] elsewhere, as in : mə^{n}gni 'asking for a girl in marriage', ḍ^{h}o^{n}g 'pretence', ba^{n}g 'highpitched

---

tied', Sanskrit bənd^{h}ən 'tie (sb.)', Persian bəndɪš 'knot'; a^{n}čəl 'one of the two ends of a sari, especially the one passing over the breast and a shoulder', Sanskrit ənčəl '(same as a^{n}čəl); region' and ančəlɪk 'regional'.

call, especially cockcrow, Muslim call to prayer', lə^{n}gṛa 'lame', rɔ^{n}gvana 'to have (something) painted', rə^{n}kta 'becomes coloured', rə^{n}gana 'to colour (tr.)'. Compare: V^{n}k [Ṽk ∞ Ṽŋ] before a voiceless consonant, V^{n}k [Ṽk] elsewhere, as in ǯ^{h}a^{n}kta 'peeps', ǯ^{h}a^{n}ka 'peeped'.

(19) V^{n}g [Vŋŋ] before a nasal; Vŋg [V̥ŋgə̆ ∞ Vŋŋ] before # or ṛ, Vŋg [Vŋg ∞ Vŋŋ] before y or v, Vŋk [V̥ŋk ∞ V̥ŋŋ] before a voiceless consonant, Vŋg [V̥ŋg] elsewhere, as in: səŋgmo^{n} 'confluences (obliqe pl.)', rɔŋg 'colour', nɪssɔŋg 'without any attachment', səŋgya 'term, noun', təŋkfʊrsət 'who lacks leisure', baŋgla 'Bengali (language)', təŋgi 'shortage'. Compare Vŋk [V̥ŋk ∞ V̥ŋŋ] before a voiceless consonant, Vŋk [V̥ŋk] elsewhere, as in: səŋkṣɪpt 'abridged', pəŋktɪ 'line (of verse)', ṭəŋkən 'typewriting', ɔŋk 'number'.

Compare the placement of the tonic in the following: [sə̥ŋ"har], [rə̃ŋ"va.nạ] (with [sə̥ŋ] and [rə̃ŋ] as medium), [nɪs"sə̥ŋŋ], ["sə̥ŋŋ.ṣɪpt] (with [sə̥ŋŋ] as heavy) from the above examples.

(20) **H** ṇ ≈ n, **U** n, mostly in Sanskrit loanwords, as: vaṇɪǯy 'commerce', gʊṇ 'quality, virtue', gʊṇa 'multiplication', ʊṣṇ 'hot', vərṇ 'letter (of the alphabet)'. (Contrast: gʊna 'times, multiplied by'.)

Note that ṇ and ŋ never occur initially and ŋ always occurs before a velar stop, m, n, or h. The only exceptions are the names of the letters ṇə and ŋə. /g/ as [ŋ] and /k/ as [ŋ] (see 18, 19) of course have greater freedom of distribution. There is, in **U**, no /ṇ/, but **U** /n/ is regularly [ṇ] before a retroflex stop. **H** /ṇ/ is [ṇ] initially, in gemination, adjacent to a retroflex stop, or before a syllable-final y or v, but [ṛ̃ ≃ ṇ] elsewhere. **H** /n/ thus never occurs before a retroflex stop or retroflex nasal. Note further that ṛ, ṛh, and (with some limitations noted in 18 above) g are fully nasalized into [ṛ̃], [ṛ̃^{h}], [ŋ] respectively after a tautosyllabic V^{n} [Ṽ], but that ŋ

[r̃], ṇ [ṇ], ŋ[ŋ], ṇg [ŋŋ], ŋg [ŋg], etc. only partially nasalize a preceding vowel (V [V̜]).

(21) The various renderings of the Sanskrit loanwords can be brought under a morphophonemic unit—ṇ before a retroflex stop; ŋ before a velar stop (further subject to 19, 20) or h (ŋh∞ ŋg^h, before # also ŋg ∞ ŋŋ); y^n ≈ ŋ ≈ n before y; v^n ≈ m before v; n before a dental stop or nasal, an affricate, a sulcate, or l; m elsewhere (the anusvāra does not occur before ṇ, ŋ, or ṛ), as in pəṇḍɪt 'Hindu scholar', səŋket 'indication, sign', səŋhar 'destruction', sɪŋh 'lion', səy^{n}yəm 'selfcontrol', səy^{n}yʊkt 'joint, united', səv^{n}vət 'year of the Vikram era', səntoṣ 'satisfaction', sənnɪpat 'fever brought on by simultaneous disturbance of the three humours (in Ayurvedic medicine)', sənskrit 'Sanskrit', sənšəy 'doubt', sənrəkṣəṇ 'guardianship, patronage', sənləgn 'connected', səmbhəv 'probable, possible', səmman 'respect', səmɪtɪ 'committee', əhəm 'ego', svəyəm 'oneself, himself, etc.'

Note that /v^n/ is, before v, [ɱ] a labiodental nasal optionally with imperfect oral occlusion and that anusvāra before a tautosyllabic sulcate is optionally /^{n}/ in some items : sə^{n}skrit for sənskrit.

(22) **H** V^n, **HU** V^n∼ V, in a few items, as : i^{n}ṭ 'brick', sɛ^{n}kṛa 'a group of hundred', ma^{n}ṭ 'earthen pot', si^{n}a 'to sew', te^{n}tis 'thirty-three', te^{n}talis 'forty-three', čo^{n}səṭ^h 'sixty-four'.

(23) **H** Vn, **U** Vn ≃ V^n, finally in some Persian loanwords, as : zəmi^n 'land', vira^n 'wilderness', pərdə+nəši^n 'behind a veil', məǰnu^n 'crazy'. The variant with V^n is especially favoured in verse and in combinations, as : zemi^n + qənd 'elephant's foot, Amorphophallus campanulatus, Blume', xa^n + sahɪb '(respectful appellative)', xu^{n}xvar 'bloodthirsty', zəmi^{n}dar 'landlord'.

A vocoid or a vocoid sequence adjacent to a tautosyllabic nasal (/m n ṇ ŋ/) is partially nasalized (symbol

a subscript hook as opposed to the superscript tilde for full nasalization). Short vowels, especially high short vowels, are less susceptible than other vocoids. A nasal preceding the vocoid in question is less influential than a nasal following : the nasalization is stronger in /kan/ 'ear' [ką̄n], /d^{h}any/ 'foodgrain' [d^{h}ąny] than in /nak/ 'nose' [nąk], /nyay/ 'justice' [nyąy̯]. Superior /^{n}/ indicates not only the full nasalization of vocoids and of l, ṛ, ṛ^h in that syllable but also the partial nasalization of any uninterrupted vocoid sequence immediately preceding a vocoid bearing ^{n}.[16]

səy^{n}yəm 'self-control' [sə̃ỹ.yəm], rate^n 'nights' [ra.tẽ], tane^n 'musical flourishes' [ta.nẽ], gʊsa^{n}i 'member of Vaishnavite monastic orders' [gʊ.sã.i], kʊ^{n}ara 'bachelor', [kʊ̃.a.ra], ba^{n}ya ~ baya^n 'left' [bã.ya ~ bą.yã], b^{h}ə^{n}vər 'whirlpool; humming bee' [b^{h}ə̃.vər].

rahe^n 'ways' [rą.ɦ̃ẽ], gaye^n 'cows' [gą.ỹẽ], matae^n, 'mothers; smallpox' [ma.tą.ẽ], b^{h}aɪo^n 'brothers (oblique pl.)' [b^{h}ą.ɪ̨.ỹõ], gəvaho^n 'witnesses (legal, oblique pl.)' gə.vą.ɦ̃õ].

kaɪa^n 'mosses' [ką.ɪ̨.ỹã], ka^{n}ɪa 'cunning' [kã.ɪ.ya]; gəvau^n 'if I had (it) sung' [gə̨.y̯ą.ũ], gə^{n}vau^n 'if I lost' [gə̃.y̯ą.ũ]. (Such contrasts—based on the number and placement of /^{n}/ in a word—are not frequent in occurrence.)

Note that a nasal cannot follow a tautosyllabic V^n, but can precede a tautosyllabic V^n—yielding a contrast between NV^n [NṼ] and NV [NV̨] in careful speech.

ma^{n}č 'rope for tying a sail', mač 'platform set up in a tree', nač 'dance'; ma^{n}g 'demand (sb.)', mag^h 'the eleventh month in the Hindu calendar'; na^{n}d 'tub', nad 'voice, sound'; tʊm^{h}e^n 'to you', kʊm^{h}ar 'potter'.

16. I am especially indebted to Dr. SUMAN (see Preface) in this paragraph.

Compare also :

siⁿna ~ sina 'to sew', sina 'bosom'; muⁿṛ 'shave' [mũr̃], guṇ 'quality, virtue' [gur̃]; raⁿṛ '(contemptuous term for) widow, prostitute' [rãr̃], baṇ 'arrow' [bąr̃]; baⁿg 'highpitched call, especially cockcrow, Muslim call to prayer' [bãŋg ∞ bãŋ], bahyaŋg ∞ bahyaŋŋ 'outward aspect' [bahyąŋg ∞ bahyąŋ·].

But in a familar style (H≈, HU≈, U≈) some of these contrasts are apt to break down. In particular a vocoid sequence followed by a tautotsyllabic nasal, or a vowel in N—C* (where C* is a voiceless consonant), or an uninterrupted vocoid sequence immediately preceding /ⁿ/ is invariably fully nasalized in a familiar style.

(24) V≈Vⁿ in such environments, as in: mač, nač, gayeⁿ, gəvauⁿ, gəⁿvauⁿ above with all vocoids fully nasalized. (The full nasalization of a in kan above cannot obliterate any contrast, since *kaⁿn is ruled out from the start. But guṇ will yield guⁿṛ in a familiar style and rhyme with muⁿṛ.)

1.2.3. The Problem of the retroflex flaps.

Initially, before a syllable-final y or v, and adjacent to another retroflex occlusive (i. e. stop or nasal), we always have the retroflex occlusives— /ḍ/, /ḍh/, and /ṇ/ [ṇ]—and never the retroflex flaps— /ṛ/ [ṛ, r̃], /ṛʰ/ [ṛʰ, r̃ʰ], and /ṇ/ [r̃].

ḍɪgna 'to be unstable', ȷ̈aḍy 'lethargy', həḍḍi 'bone', əṇḍa 'egg', brɛṇḍi ~ braṇḍi 'brandy'; ḍʰəŋg 'manner', sənaḍʰy '(name of a caste)', čəḍḍʰa (surname)', ṣəṇḍʰ 'sexually impotent, coward'; kəṇv '(name of a ṛṣī, the foster-father of Śakuntalā)', əkṣuṇṇ 'not beaten (track)'. (The name of the letter ṇ is : ṇə [ə̆"r̃ə≃"ṇə].)

Again, a retroflex flap never occurs if r follows with or without an intervening vocoid sequence; a retroflex

occlusive can occur in this environment, as in: gəḍərɪa '(name of a shepherd caste)'.

Intervocalically, finally after a vowel, and elsewhere, the flaps are more frequent than the stops are. In these environments, /ṇ/ is [r̃ ≃ ṇ] (examples under 20 above). (U [ṇ] is of course simply a variant of U/n/ before a retroflex stop.) With the stops, we even encounter a few cases of positional alternations like əṛʰai ~ ḍʰai 'two-and-a-half', buṛʰa ~ buḍḍʰa 'old'. Examples of ṛ and ṛh are :

gaṛi 'cart', guṛ 'coarse brown cane sugar', ləṛka 'boy', səṛkeⁿ 'roads', čəmṛa 'leather', həṛtal 'strike, hartal', uṛd 'black gram'; siṛʰi 'ladder', ḍeṛʰ 'one-and-a-half', pəṛʰna 'to read'. (The names of the letters are : Hindi /ṛə/ [ə̆''ṛə], /ṛʰə/ [ə̆''ṛʰə], and Urdu /əṛe/ [ə''ṛe] ~ /ṛe/ [ə̆''ṛe].)

There are clear counter-examples, however, which force us to recognize /ṛ/ and /ṛʰ/ as phonemes distinct from /ḍ/ and /ḍʰ/ respectively. The contrasts of the following type leave us no other choice :

əḍɪg 'firm', nə+ḍɪgna 'not to be unstable', həḍḍi 'bone', gɪr+pəṛi 'fell down'; nɪḍʰal 'wearied', nə+ḍʰalna 'not to spill (tr.)', čəḍḍʰa '(surname)', čɪṛʰa 'was angered; anger (vb. imperative)'.

The plus juncture is of no help, though it must certainly be noted in passing that most cases of /ḍ/ and /ḍʰ/ in anomalous environments coming under 25, 26, and 27 below are cases of intervocalic stops preceded by a morpheme cut in which the latter morpheme is one that occurs elsewhere after #.

əḍɪg 'firm', nɪḍər 'fearless', suḍol 'graceful', beḍʰəṅga 'clumsy' by the side of ḍɪgna 'to be unstable', ḍər 'fear (sb., vb. imperative)', ḍol 'manner', ḍʰəṅg 'manner'; ḍugḍugi '(kind of small drum)', ḍəbḍəbana 'to fill with tears'.

Most, but not all.

baḍi ~ baṛi 'small enclosed garden', əḍana ~ əṛana '(name of a *raga*)', laḍ ~ laṛ 'expression of fondness, love', ḍɪḍkari ~ ḍɪṛkari 'sudden outcry', nɪḍ^{h}al 'wearied', roḍa '(surname)' (presumably to avoid homonymy with roṛa 'brickbat'!), uɟ̃əḍ ~ uɟ̃əṛ ~ uɟ̃əḍḍ 'fool'.

(25) H ṛ ≃ ḍ, HU, U ṛ in /ṛ/-preferring environments in Sanskrit loanwords, as: aḍəmbər 'ostentation without substance', təḍɪt 'lightening', uḍu 'star', k^{h}əḍg 'sword'. Similarly with H ṛ^{h} ≃ ḍ^{h}, HU, U ṛ^{h}, as in: guḍ^{h}art^{h} 'occult meaning'.

(26) ḍ ≈ ṛ in /ṛ/-preferring environments in English loanwords, as: soḍa 'soda water; sodium (bi)carbonate', pəgoḍa 'pagoda' pɛḍ 'writing pad', bʌḍɪs ~ bʌḍi 'Western-style bodice', muḍ 'mood (psychological)'.

(27) H ḍ [^{n}ḍ] in V^{n}—, HU ḍ ~ ṛ in V^{n}—, U ṛ [r̃] in V^{n}—#, V^{n}—C* but ḍ [^{n}ḍ] in V^{n}— elsewhere (where C* is any nonsyllabic other than a retroflex occlusive, r, a syllable-final y or v, or h), as in: go^{n}ḍ 'Gond (a tribe in Central India)', mi^{n}ḍ 'ornamental glide (in Hindustani music)', mʊ^{n}ḍta 'shaves'; go^{n}ḍa 'water-tank', mʊ^{n}ḍa 'shaved', ɪ^{n}ḍuri ~ i^{n}ḍuri 'soft cloth pad used in carrying a load over one's head or resting certain types of drum on the ground', gə^{n}ḍeli 'piece cut off from a sugarcane stem', mə^{n}ḍrana 'to hover around'. Similarly with the corresponding aspirates: ḍ^{h}u^{n}ḍ^{h}ta 'looks for'; ḍ^{h}u^{n}ḍ^{h}a 'looked for', mu^{n}ḍ^{h}a '(kind of cane seat)', ḍ^{h}ə^{n}ḍ^{h}ora 'drum used by a public crier', ḍ^{h}u^{n}ḍ^{h}aṛi '(name of the Rajasthani dialect spoken around Jaipur)'.

There is a good deal of vacillation in assigning a lexical item to V^{n}ḍV or V^{n}ṛV (where ṛ is [ṛ], not [r̃], since V^{n} belongs to a different syllable). Consequently a given idiolect may come up with some contrasts (cf. 15 above). Thus:

lɔ^{n}ḍa '(slang) boy', lɔ^{n}ḍa ~ lɔ^{n}ṛa 'boy partner in

paederasty' (cf. lɔṛa '(vulgar) penis'); lɔ^{n}ḍi '(slang) girl', lɔ^{n}ḍi ~ lɔ^{n}ṛi 'female slave'.

We can also mention in this connection a few sporadic cases of alternation : ṛ ~ r, l ~ r (with **U** registering a slight preference for r in either case) : ɣʊbbaṛa 'balloon', phʊlvaṛi 'small enclosed flower garden', kəṛoṛ 'crore' (the first ṛ, see also 73), əṛaṛ 'cattle shelter' (the first ṛ), nɪvaṛ 'broad cotton tape', kəčɔṛi '(kind of pie with salty filling)', ʊṛd 'black gram', sɔṛsəṭ^{h} '67', rəbəṛ 'india-rubber', məčč^{h}əṛ 'gnat, mosquito', saṛi 'sari', puṛi 'deep-fried wheatflour cake', bɔč^{h}aṛ 'shower'; gə^{n}ḍeli 'piece cut off from a sugarcane stem', mə^{n}ḍlana 'to hover around', sʊnhəla 'golden' (see 50 below), sɔlhəǯ 'wife's brother's wife'.

1.2.4. The Problems of the glottal fricatives.

The visarga in Sanskrit loanwords simply geminates a following sulcate and is no problem, for example, in :

nɪssəng 'without any attachment', nɪššʊlk 'free of charge'.

(28) **H** h ≃ hh, **HU**, **U** h, corresponding to a visarga in V—#, as in : pratəhh 'in the morning', šənɛhh 'slowly'.

(29) **H** hP*$_1$ ≃ hP*$_1$P*$_1$, **HU**, **U** hP*$_1$ (where P* is k, k^{h}, p, p^{h}) corresponding to a visarga in V—P*, as in pratəhkkal 'morning', əṇtəhppʊr 'seraglio'.

Phonetically hh in V—# is [ḫ], h in V—P*$_1$P*$_1$ is [ḥ]. Note the placement of the tonic in "pra.təh ≃ pra"təhh, pra.təh"kal ≃ pra"təhk.kal.

Comparable is the intervocalic geminate h.

(30) **H** h, **HU** h ~ hh, **U** hh occurring intervocalically as in : ohho '(exclamation of exasperation)', sɪhhɔt 'health' (**H** sehət).

Phonetically hh in V—V is [ḫh].

From now on we shall treat aspiration (written as a superior h) as a separate phoneme much like /^{n}/ and different from /h/. This /^{h}/ occurs only after p, t, ṭ, č, k, b, d, ḍ, ǯ, g, m, n, l, r. ṛ. It would be nice, if we could treat it as a positional variant of /h/ occurring after C* (where C* is p, t, etc.) but not after ŋ, s, h, v, etc. Compare :

səŋhar 'destruction', əshay 'helpless', məšhur 'famous', məzhəb 'religion, faith', sɪhhət 'health', əyhɪk 'of this world', əvhelna 'unfeeling neglect'.

Treating [p^{h}], [t^{h}], etc. as aspirate clusters in this way would enable us to make better sense of the similar allophonic treatment of syllable-final [^{h}] after voiced consonants and /h/ after vowels in **HU**, of their relative lack of freedom of distribution (unlike most other nonsyllabics, for example, they do not geminate either intervocally or finally after a vowel), and of certain alternations, as : bəhɪšti ≈ b^{h}ɪšti 'water carrier', məharaǯ ≈ m^{h}araǯ 'sir' (110);[17] səŋhar ∞ səŋg^{h}ar 'destruction' (21); səb + hi ∞ səb^{h}i 'all (emphatic)' (sporadic, cf. fn. 14 end); the Rules v, vi, vii as they variously apply to (49) through (54); and the following (31, 32, 33).

(31) C*^{n}C*^{h} ∞ C*C*^{h} (the former is used only in very slow tempo), as in : me^{n}ǯ^{h}d^{h}ar 'midstream', (∞ mə^{n}ǯd^{h}ar) əd^{h}k^{h}ʊla 'half-open', kəṭ^{h}p^{h}oṛva ~ keṭ^{h}p^{h}ʊṛva 'woodpecker'. Similarly : C*^{h}h ∞ C*h, hC*h ∞ C*h, as in : dʊk^{h}hari ~ dʊk^{h}yari 'pain-killing' (∞ dʊkhari), ərt^{h}hin 'impoverished; meaningless', əṭ^{h}həttər 'seventy-eight';

---

17. The aspirates m^{h}, n^{h}, l^{h} do not ordinarily occur initially; if they do, it is only as morphophonemic alternants (33, 110). Ordinarily r^{h} does not occur in any position; morphophonemically it does, however (33, 34, 47-54, 110). If a sequence of the type C*r^{h}, C*l^{h} where C* is a stop turns up, it is always to be rewritten as C*^{h}r, C*^{h}l respectively. Thus brəhm (see 52) becomes not *br^{h}ɛm but b^{h}rɛm.

mʊh^{n}b^{h}ərai 'bribe paid for sealing someone's mouth' (∞ mʊ^{n}b^{h}ərai).

(32) If a word contains within itself two glottal fricatives (/h/ and /^{h}/) not forming sequences dealt with in 30, 31, then the later of the two is dropped in familiar or rapid styles, as in : ṭ^{h}əṭṭ^{h}a 'mocking' (≈ ṭ^{h}əṭṭa ∞ ṭ^{h}əṭṭa), ṭ^{h}əṇḍ^{h}ək 'coolness', č^{h}ɪasəṭ^{h} 'sixty-six', b^{h}uk^{h} 'hunger', ǯ^{h}uṭ^{h}a 'false', b^{h}ab^{h}i 'brother's wife', p^{h}ʊpp^{h}ʊs 'lung'; p^{h}ʊhara 'spray' (≈ p^{h}ʊara ∞ p^{h}ʊara), č^{h}ɪhəttər 'seventy-six', b^{h}ɔh^{n} 'eyebrow'; grɨhəst^{h}i 'household' (≈ grɨhəsti ∞ grɨhəsti), səhpaṭ^{h}i 'classmate'; səhih 'correct' (≈ səhi ∞ səhi). Note, however; dʊkk^{h} 'pain' (and not *dʊhkk(^{h}) according to 29 and 32).

(33) C*V̄h ≈ C*^{h}V̄ (where h is tautosyllabic to the long vowel), as in; gʊnah 'sin, crime' (≈ gʊn^{h}a), mahvari 'monthly', nɪbahna 'to manage to preserve or complete' sərahna 'praise, to praise', kərahna 'to groan', təfrih 'pastime, pleasure walk'. Contrast : kəraha 'groaned' (and not *kər^{h}aya), gʊnaho^{n} 'sins, crimes (oblique pl.)' (and not *gʊn^{h}ao^{n}).

We already envisage the possibility of C*h distinct from C*^{h} in 31 above. There may or may not be any morpheme cut between C* and h. Examples follow which clearly show that /h/ and /^{h}/ are distinct phonemically.

ʊphar 'present, gift', pəṭhar 'one whose occupation is threading ornaments and lacework', səth-ttər 'seventy-seven', ɪkhəttər 'seventy-one', ənhoni 'untoward (event)', honhar 'promising', fɪlhal 'for the moment', ləkəṛhara 'wood-cutter' (all with a morpheme cut before h).

fəthə 'zabar, the diacritic for /ə/', səthi 'superficial'. nɪkhət 'pleasant smell', gədha 'donkey', əghən 'the ninth month in the Hindu calendar', ləmha 'moment',

nənha ‘small child’, mənhus ‘ill-omened’, ɪnhɪraf ‘disobedience’, dulha ‘bridegroom’, ɪlham ‘intuition’, pərhez ‘abstinence’, ərhət ‘(a class of Jain divinities)’, puǰarh ‘fit to be worshipped’.

(34) $\breve{V}C^*h\bar{V} \approx \breve{V}C^{*h}\bar{V}$, as in: gədha, pərhez, etc. above.

Note the placement of the tonic in "gəd.ha ~ gə"dʰa.

Other alternations involving /h/ and /ʰ/ are described in § 1.2.5 since they involve vowels also.

1.2.5. The Problems of an enlarged vowel inventory.

At the phonetic level at least, the core system is found to be overly simple. To start with we need a larger inventory of vowel phones (see Table 3).

TABLE 3

i	ĭ			u
	ɪ	ɨ	ʊ	
e				o
	ĕ	ə̆	ŏ	
ɛ	ɛ̆	ə	ɔ̆	ɔ
		ă		
		a		ᴀ

[ɨ ĕ ŏ] are, like [ɪ ʊ ə], short and lax. [ɛ̆ ɔ̆ ă] are short and tense and somewhat more neutralized than the short lax group. [ɛ ɔ ᴀ] are, like [a], long and lax. [i u e o] are, of course, long and tense. [ɨ] is, like [ɪ ʊ], lower high. [ĕ ə̆ ŏ] are mid or slightly higher. [ᴀ] is low and back and may optionally be slightly rounded. [ĭ ə̆] are lax and phonetically syllabic, but being extra-short and predictable, they do not count as vowels phonemically—unlike all the rest they do not count in the placement of the tonic and do not participate in the various positional variations of vowel phonemes described from time to time (in § 1.1.2 and later). [ɪ ʊ] are optionally

weakly voiced when final and at the same time with the penultimate syllable medium or heavy.

This inventory with the foregoing phonetic notes will enable us to simplify the following description considerably. Some of the problems involving additions to the core inventory will be taken up later : [ɛ ɔ] and the diphthongs in § 1.2.6 and [ĭ ə̆] in § 1.2.7.

Final vowels are generally peripheral or long. But there are exceptions—especially in C—#.

(35) **H** ɪ, **U** i occurring in C—# mostly in Sanskrit loanwords of more than one syllable, as : lɪpɪ 'script', ʊtpəttɪ 'creation', hanɪ 'harm, loss', b^{h}a^{n}tɪ 'manner, way', šantɪ 'peace'.

(36) **H** ʊ, **U** u similarly to 35, as in : gʊrʊ 'guru, mentor', bɪndʊ 'point', b^{h}anʊ 'sun'.

(37) **H** ɪ, **U** ɪ [ĕ] occurring finally in monosyllables, as : kɪ '(conjunction)' (and its combinations, as : bəl + kɪ 'on the contrary', ta+kɪ 'so that', čʊ^{n} + kɪ 'because'). (Contrast : ki 'of, 's').

(38) **H** o, **U** o ≈ ʊ [ŏ] occurring finally in monosyllables, as : ko 'to'.

(39) **H** e, **U** ɪ [ĕ] occurring in Persian loanwords of more than one syllable, as : əgərčɪ 'although'; beda^{ʔ}ɪ '(pl.) wonders'; bəǰaɪ + xud 'in place of oneself', pərdɔɪ + simi^{n} 'silver screen', činɪ + ǰəbi^{n} 'frown', dərdɪ + dɪl 'heartache' (and other instances of the Persian enclitic {izāfat} 'of'). (See also 73 below.) (When followed by close juncture, {izāfat} is **H, U** ɪ, as in kohɪnur 'mountain of brilliance (name of the famous diamond)').

(40) **H** o, **U** ʊ [ŏ] similarly to 39, as in : təvəqqʊ 'expectation'; šo^{ʔ}ləʊ + šəbnəm 'flame and dew', vəzi^{ʔ}ʊ + šərif 'the bad and the good (people)', abʊ + həva 'water and air, climate', tərʊ + tazə 'moist and fresh, feeling fresh',

(and other instances of the enclitic form ʊ of the Persian connective {və} 'and'); əllahʊ+əkbar '(the takbīr proclaiming that) God is great'.

H ɪ ʊ do not occur in V—#, elsewhere finally they are : [ɪ ʊ]. U ɪ ʊ finally are [ĕ ŏ]. Needless to say, [ĕ ŏ] do not occur finally in H and [ɪ ʊ] do not occur finally in U.[18]

(41) [ə̆], H ≃ ə occurring finally after some clusters and after the sequences iy ɪy ʊv ey ɛy (these latter occurring only in Sanskrit loanwords) as in : mənd 'quiescent, slow', čənd 'a few', sʊst 'lethargic', gošt 'meat', bʰədiy 'yours faithfully', prɪy 'dear', dʰrʊv 'Pole Star', pey 'drink'. (For further details see § 1.2.6–7; ɪv, uy, etc. end in a semivowel phonetically.)

(42) ə occurring finally in monosyllables, as : nə 'not', kə 'the letter *k*', bə+xʊši 'with pleasure'.

(43) H, HU a, U ə[ă] occurring in C—# in some Persian loanwords of more than one syllable (corresponding to the "silent" he and to 'ain in the orthography) as : pərvanə 'moth; licence', sadə 'plain, ordinary', avarə 'vagrant', mevə 'fruit', bəččə 'young one, child', bəndə 'servant', məlikə 'queen', vərnə 'otherwise', ahɪstə 'slowly', faxtə 'falcon', fəthə 'zabar, the diacritic for /ə/', šoʔlə 'flame' (all ending in silent he); ǰamɪʔə 'university' (with 'ain followed by silent he); vəzʔə 'manner', zɪlʔə 'rib; district' šəmʔə 'candle' (the last three ending in 'ain).

H ə is [ə] finally and occurs except in H ≃ only in

18. In the sorting out of many alternations in this section (37–40, 44–54), I was greatly helped by Dr. Jain (see Preface), and the dictionaries of Platts and Maddāh (J. T. Platts, *A Dictionary of Urdū, classical Hindī, and English*, 4th printing, London : Crosby Lockwood & Son, 1911 and M. Mustafa Khan Maddah, *Urdū-hindī śabdakoś*, Lucknow : Government of U. P., Department of Information, 1959).

monosyllables. **HU** ə is [ə ≈ ɛ̆] finally and occurs only in monosyllables. **U** ə is [ă] finally—as in nə, pərvanə above (42, 43) or in əšokə 'Emperor Aśoka' or in əšokə + hoṭəl (but əšok as a modern male given name). Final [ə̆] is predicatable and phonemically nothing. Final [ə̆] versus [ə] makes a difference to the placement of the tonic, as in : gəṇ''təntr, **H** ≃ ''gəṇtəntrə 'republic'. The same is true of final [ă] versus [a], as in : **U** ''avarə, **H**, **HU** a''vara above.

Before dealing with sequences of vowel followed by h, it is useful to recall that h is followed by a vocalic off-glide echoing the preceding vowel in V—C and, in **H** only, in V—#; and that [Vh] in the same syllable is [V_hh] or, **HU** ∞ [V_h] where [V_h] stands for a breathy vowel.

U ɪ ʊ are respectively [ĕ ŏ] in the following environments also : adjacent to q or ˀ, before a tautosyllabic h, and, in **U** ≃ only, before a heterosyllabic h.

(44) **H** ɪ ≈ e, **U** ɪ occurring adjacent to q, after ˀ, or before a heterosyllabic h mostly in Persian loanwords, as in : qɪblə 'the holiest spot in Mecca', ɪqrar 'contract', ˀɪvəz 'substitute', ˀɪlm 'knowledge', šaˀɪr 'poet', sanɪhə 'calamity', ǯɪhad 'holy war', rɪha 'free'. Similarly with **H** ʊ ≈ o, **U** ʊ, as in : qʊtb 'Pole star', qʊrˀan 'the Qur'ān', hʊqqə 'hookah', ˀʊmr 'age', ˀʊsman '(Muslim male given name)', sʊhag 'conjugal bliss', mʊhaǯɪr 'refugee', mʊhavərə 'idiom'.

(45) **H** e, **U** ɪ occurring before ˀ or a tautosyllabic h mostly in Persian loanwords, as in : fɪˀl 'deed, verb', ɪˀlan 'declaration', vaqɪˀətən 'real, true', ɪstɪˀmal 'use', yɪh 'this', sɪhhət 'health', mɪhman 'guest', mɪhⁿdi 'henna, Lawsonia inermis, L.', mɪhr 'sun; affection', ɪhsan 'favour', bɪhterin 'best', sɪhra 'ceremonial headdress for the bridal couple', čɪhra 'face'. Similarly with **H** o, **U** ʊ, as in : šʊˀəra 'poets (collectively)', vʊh 'that, he.

she, it', kʊhra 'mist', mʊhr 'mohur (a formerly used gold coin)', mʊhre^{n} 'mohurs', sʊhbət 'companionship', kʊhl 'kohl (for darkening eyelids)'.

(46) **H** ɪ, **U** ɪ before a tautosyllabic h, as in: ɪhlok 'this world', tɪhra 'threefold; earthen pot'. Similarly with **H** ʊ, **U** ʊ, as in: mʊh^{n} 'mouth, face', kʊhro^{n} 'holes, caves (oblique pl.)'. Compare also **H** e, **U** e before h, as in gehu^{n} 'wheat', dehat ~ deh 'village', sneh 'friendship', aramdeh 'comfortable', snehšil 'friendly'. Similarly with **H** o, **U** o, as in: tyohar 'festival', mohən '(male given name)', koh 'mountain', gʊroh 'crowd', tohfa 'present, gift'.

With alternations involving əh the situation is more complicated. The sequence əhʊ followed by C is **H** [əhʊ], **HU** [əhʊ ∞ ɔ̃h], **U** [ɔ̃h].

(47) əhʊ, **HU** ≈ ^{h}ɔ (subject to Rule viii below) occurring in —C, as in: bəhʊt 'many, very', pəhʊ^{n}čna 'to reach', bəhʊtere 'several', pəhʊ^{n}čana 'to cause to reach'.

All other cases of ə followed by a heterosyllabic h remain əh [əh], as in:

məhʊa 'Madhuca indica, J. F. Gmel. and M. longifolia, (Koenig) Macbr.', bəhu 'bride', vəhi 'that very, he alone, etc.', vəha^{n} 'there', rəha 'stayed', pəheli 'puzzle', bəhɪn ~ bəhən 'sister', rəhe + ga 'will stay (masc.)'

Cases of ə followed by a tautosyllabic h may be best presented in a series of compact rewrite rules starting with the morphophonemic spellings //h// in C—#, //hə// in C—C, //əhə// in —C, //əh// in —#, —C, —v, //ə*h//. These rules will also involve the phonemic sequence əhə with or without the free variant əhɪ. They are to be applied in order.

(i) h in C—# → əh, **U** ≃ h

(ii) hə in C—C → hə ≈ əh

(iii) əhə in —C → əhə ∞ əh

(iv) əh in —C # → əh ≈ əhə

(v) əh in —C, —v → əh ≈ ʰɛ

(vi) əh in —# → əh ≈ ʰɛ ≈ ʰe

(vii) ə*h → əh ≈ ʰa ≈ a

(viii) ʰ not after C* (where C* is a voiced "aspirable" consonant—b, d, ḍ, ǯ, g, m, n, l, r, ṛ) → ⊙ (See also fn. 17.)

(ix) əh not followed by a vowel → **H** [əh], **HU**, **U** [əh ∼ ɛ̆h] syllable-finally, [ɛ̆h] elsewhere. Examples: rəh 'stay (imperative)', rəhta 'stays', rəhn 'mode of living'. (Compare the phonetic rule above about əhʊ (variously [əhʊ] and [ɔ̆h].)

Note that [ɛ̆ ɔ̆] cannot be identified with the phonemes /ɛ ɔ/ (as has been done before, by GUMPERZ and NAIM (fn. 2) for example). First, [ɛ ɔ] too can occur before a tautosyllabic h, as in:

nɛhroⁿ '(oblique plural of nɛhər, married woman's parental home)', gɔhroⁿ '(oblique plural of gɔhər, pearl)', ǯɔhəri ∞ ǯɔhri 'jeweller'.

Secondly, [ɛ̆ ɔ̆] are short for the placement of the tonic unlike [ɛ ɔ]. Note the placement of the tonic in: məhtab 'moon' [mɛ̆h"tab], rəhta ≈ rʰɛta 'stays' ["rɛ̆h.ta ≈ "rʰɛ.ta], bəhʊtere ≈ bʰɔtere 'several' ['bə.hʊ"te.re ∼ bɔ̆h"te.re ≈ bʰɔ"te.re], bəhʊtoⁿ ≈ bʰɔtoⁿ 'many (oblique pl.)' ['bə.hʊ"tõ ∼ " bɔ̆h. tõ ≈ "bʰɔ.tõ].

Thirdly, [ɛ̆ ɔ̆] never alternate with the diphthongs as [ɛ ɔ] do.

(48) **H** əhɪ, **HU**, **U** əhə before a tautosyllabic C (subject to iii, iv, viii, ix) but əh before a heterosyllabic C (subject to v, viii, ix), as in: bəhɪn 'sister', pəhɪla 'first', pəhɪle 'at first'. (With pəhčan 'acquaintance', gəhna 'ornament; pawned article', pəhənna 'to wear',

etc. the forms with əhɪ are now near-obsolete...Note also pəhnana ≈ pɪnʰana 'to cause to wear'.)

(49) əh, U ≃ h [hə̆] occurring in C – # (subject to i, vi, viii, ix) in Persian loanwords, as in ɪ vəʒh 'reason', tərh 'kind', fəth 'victory', səth 'surface', mədh 'praise'.

(50) hə in C–C (subject to ii, iv, v, viii, ix), as in: ərhər 'pigeon pea, Cajanus cajan, Mill.', kəṭhəl 'jack fruit, Artocarpus integra, Thumb.', sʊnhəla ~ sʊnhəra 'golden', dʊlhən 'bride', dəšhəra 'the ten days ending in the festival of Vijaya Dashami, Dassehra'.

(51) əhə in –C (subject to iii, iv, v, viii, ix) mostly in Sanskrit and Persian loanwords, as: ləhər 'wave', məhətv 'importance', səhər 'morning', rəhəm 'mercy', məhəbbət ~ mʊhəbbət 'affection', pəhɪle + pəhəl 'at the very outset', rəhən 'mode of living', rəhən + səhən 'ditto'.

(52) əh in –C, –V (subject to iv, v, viii, ix), as in šəhr 'city', nəhr 'canal', məhl 'palace', brəhm 'Brahma, the cosmic principle', məhkna 'to smell sweetly', pəhlvan 'wrestler', pəhn + lo 'put (it) on', ṭʰəhrna 'to stay', pʰəhrna '(of flag) to fly', kəčəhri 'office', pəhčan 'acquaintance', gəhna '(silver, gold) ornament', səhvas 'living with, company', məhⁿga 'costly'. ləhⁿda 'Lahnda, spoken in W. Panjab', ləhreⁿ 'waves', mehloⁿ 'palaces (oblique pl.)', rəhta 'stays'.

(53) əh in – # (subject to vi, viii, ix) as in: rəh 'stay (imperative)', kəh 'say (imperative)', čʰəh 'six',[19] təh 'fold, layer', məh 'moon'; pratəhh 'in the morning' (see 28).

(54) əh ≈ ʰa ≈ a (subject to vii, viii, ix but not to iv, v, vi) occurring after C in a few items, as: ʒegəh 'place', sʊbəh 'morning', məsaləh 'spice; ingredient', ʒəgəheⁿ 'places'.

19. In Devanagari, the numeral six is often spelt as *chaḥ* (like *prātaḥ*). It will make more sense though to spell it as *chaha* (like *kaha*).

We have been successful so far in leaving the inventory of vowels phonemes ɪ ʊ ə, i e u o a intact and at the same time "taking care of" [ĕ ŏ ɛ̆ ɔ̆ ă] by putting [ĕ] under ɪ, [ŏ] under ʊ, and [ɛ̆ ɔ̆ ă] under ə. It is possible, however, that in a given idiolect sporadic contrasts—say, between [ɪ] and [ĕ], between [ʊ] and [ŏ], or between [ə] and [ɛ̆]—may crop up due to code-switching and bidialectal interference.[20]

An enlargement of the phonemic inventory is unavoidable in the following cases of loanword vowels :

(55) ĕ ≈ e occurring in English loanwords, as : čĕk 'cheque', čĕkoⁿ 'cheques (oblique plural)', fᴀrĕn 'abroad'.

(56) ᴀ ≈ a occurring in English loanwords, as : bᴀl 'ball'. kᴀfi 'coffee', lᴀ + kᴀlĕǯ 'college of law'. (Note that these words never show /ɔ/ in place of /ᴀ/. Compare 59 below.)

(57) H rɪ ≃ rɨ, HU rə ∼ ɪr, U rə ≈ ər (Vr alternants only before a heterosyllabic C) occurring in Sanskrit loanwords, as : hridəy 'heart', krisṇ '(male given name, Krishna)', sənskrɨt 'Sanskrit', krɨpa 'favour, mercy', pritʰvi 'earth', rɨṣi 'ṛṣi', grɨh 'house', grɨhṇi 'housewife', əmrɨt 'immortalizing drink (in Hindu myth)'. Similarly with H lɪ ≃ lɨ, etc., as in : klɨptɪ 'stratagem'.

(58) rə, H ≈ (rɨ ∼ rɪ), occurring in hypercorrect forms : grəh 'planet', brəǯ 'Braj dialect' (H ≈ (vrɨǯ ∼ vrɪǯ)).

1.2.6. The Problem of the semivowels.

The positional variants of the phonemes /y/ and /v/ are primarily conditioned by the tautosyllabic adjacent vowel. /y/ has three variants : [i̯] before u, ʊ, e, o or after u; [ɪ̯] before ə, ɔ, [ə̆] or after e, o, ə, ɔ; and [e̯] before

20. For the irregular alternations affecting the proximal and the distal/unmarked pronominal stems, see § 1.2.9.

a or after a. Correspondingly /v/ has three resonance qualities : [u̯] before i, ɪ, e, o or after i, ɪ; [ʊ̯] before ε, ə, [ə̆] or after e, ε, ə; [o̯] before a or after a. /v/ in addition may optionally have consonantal constriction with slight friction at the lips, especially before a tonic syllabic or before [ə̆]. This constriction is bilabial before o, labio-dental elsewhere. [i̯] and [u̯] represent somewhat tenser and more fricativized vocoids in **U** than in **H**. [e̯] and [o̯] represent somewhat lower than cardinal vocoids in **U** but near-cardinal in **H**. [y] will be used as a cover symbol for [i̯], [ɪ̯], [e̯]. [w] will be used as a cover symbol for [u̯], [ʊ̯], [o̯] without [β] or [v] coarticulation.[21]

These nonsyllabic vocoids are to be clearly distinguished from the diphthongal glides. /ε/ and /ɔ/ are both long and lax; phonetically /ε/ is **H** [əe̯], **HU** [εe̯], **U** [ε] and correspondingly /ɔ/ is **H** [əo̯], **HU** [ɔo̯], **U** [ɔ]. For the placement of the tonic /ε/ and /ɔ/ always count as long. The contrast between /əy/ [əɪ̯] and /ε/ and that between /əv/ [əʊ̯] ~ [əʊ̯$_v$] and /ɔ/ are not firmly established.

(59) ε (never əy), occurring in English loanwords, as : bεṭ 'bat (for playing)', bεŋk 'bank (commercial)'. (Compare 56 above. Note at this point brεṇḍi ~ braṇḍi 'brandy'.)

(60) əy (never ε) before a vowel, as in : gəya 'went', čəyən 'collection', bəyan 'conversation, lecture'. Compare : rεˀan 'adolescence', tεya 'melted', nεyaɪk 'one well versed in Nyāya philosophy'.

(61) əv (never ɔ) before a vowel, as in : təva 'hot plate', səva 'one and a quarter', $\mathrm{b^h}$əvən 'house', gə^{n}var 'rustic', əvəst^{h}a 'age', həva 'air', həvəs 'greed, fondness.' Compare : yɔvən 'youth', sɔva^n 'hundredth', nɔva^n ~ nəva^n 'ninth'.

---

21. I am especially indebted to Mr. Dɪxɪᴛ (see Preface) in this paragraph.

(62) əyy HU ≈ εy occurring in –V, as in : səyⁿyəm 'self-control'; bəǯvəyya 'instrumentalist', rəvəyya 'custom' (U ≃ rəviə), təyyar 'ready', mʊˀəyyən 'decided upon', səyyɪd '(name of a clan, Muslim title of respect)'. Compare : nyayy 'just (demand, solution, etc.)'.

(63) H əɪ, HU əyy, U əɪ occurring in –V in a few items, as : məɪa 'mother', bʰəɪa 'brother', rəɪət 'the ruled'.

(64) əvv, HU ≈ əv occurring in –V as in : səvⁿvət 'year of the Vikram era', həvva 'bogey', fəvvarə 'fountain', qəvval 'singer of the qavvālī', čəvvalis 'forty-four'. Compare : qʊvvət ≈ quvət 'power'.

(65) H əʊ, HU əvv, U əʊ occurring in –V in a few items, as : kəʊa 'crow', čəʊən 'fifty-four'.

Note the placement of the tonic in rə''vəy.ya, 'mə.ɪ''a, ''həv.va, 'kə.ʊ''a above.

(66) H əy, HU ε, U əy occurring in environments other than –V, –y in Sanskrit and Persian loanwords, as: ǯəy 'hale, victory', vɪnəy 'modesty', bʰəy 'fear', təy 'decision', qəy 'vomiting', əy 'O (voc.)', məy 'wine'; əyhɪk 'of this world', əytɪhasɪk 'historical'; vɪnəyšil modest', ǯəyǯəyvənti '(name of a *rāga* in Hindustani Music)'. (All Sanskrit loanwords with final əy or initial əy not followed by V or y and all Persian loanwords with final əy belong here.) Compare : H hɪmaləy, U hɪmalyə 'the Himalayas'.

(67) H əv, HU ɔ, U əv occurring in environments other than –V, –v in Sanskrit and Persian loanwords, as : madʰəv '(male given name)', gɪrəv 'pawning'; əvtar, 'avatar, incarnation', əvhelna 'unfeeling neglect'; nəvyʊvək 'young man' (cf. 94), čəvkidar 'watchman'. (All Sanskrit loanwords with final əv or initial əv not followed by V or v and all Persian loanwords with final əv belong here.)

(68) **H** ɛ [əe̯] ~ əy [əɪ̯], **HU** ɛ [ɛe̯], **U** ɛ [ɛ], as in : hɛ 'is', pɛsa 'money, 1/64 (now 1/100) of a rupee', bɛl 'bull', ɛsa 'such', sədɛv 'usually', nɛhər 'married woman's parental home', vɛšnəv 'devotee of Vishnu, Vaishnavite', nɛyaɪk 'one well versed in Nyāya philosophy', tɛya 'melted', rɛʔan 'adolescence', ɛš 'luxury', ɣɛr+hazɪr 'absent'. (Compare 59 and 66).

(69) **H** ɔ[əo̯] ~ əv [əu̯], **HU** ɔ[ɔo̯], **U** ɔ[ɔ], as in : sɔ 'hundred', sɔtela 'step–, half–(brother, etc.)', kɔn 'who ?', ɔr 'and', ǯɔhər 'jewelry; self-immolation of Rajput women', kɔmary 'adolescence', yɔ̀vən 'youth', sɔva^{n} 'hundredth', ʔɔrət 'woman', ɔsət 'average', nɔǯəvan 'young man', fɔǯ 'army'. (Compare 56 and 67.)

The contrast between /e^{n}/ and /ɛ^{n}/ and that between /o^{n}/ and /ɔ^{n}/ are not firmly established. (Vowels accompanied by /^{n}/, it will be recalled, are slightly higher than elsewhere.)

(70) **H** e^{n}, **HU** ɛ^{n}, **U** e^{n}, as in : se^{n}k 'applying heat by roasting, fomenting', me^{n}ḍ^{h}ɔk 'frog', ge^{n}d 'ball', ge^{n}ḍa 'rhinoceros', b^{h}e^{n}ga 'squint-eyed', rate^{n} 'nights', me^{n} 'in', eh^{n} ~ he^{n} '(exclamation of surprise)'. Similarly with : **H** o^{n}, **HU** ɔ^{n}, **U** o^{n}, as in : g^{h}o^{n}sla 'nest', ləṛko^{n} 'boys (oblique pl.)', b^{h}o^{n}du 'fool', ho^{n} 'become, were (contingent 1st and 3rd pl.)', po^{n}č^{h}na 'to wipe', ǯo^{n}k 'leech'.

(71) ɛ^{n}, as in : b^{h}ɛ^{n}s 'buffalo', sɛ^{n}talis 'forty-seven', mɛ^{n} 'I', hɛ^{n} 'are', bɛ^{n}gən 'brinjal, egg-plant', ɛ^{n}ṭ^{h}ən 'twist : assuming airs by withholding'. Similarly, with : ɔ^{n}, as in : b^{h}ɔh^{n} 'eyebrow', lɔ^{n}g 'clove', b^{h}ɔ^{n}kna 'to bark', čɔ^{n}ka 'was startled', sɔ^{n}f 'aniseed, Pimpinella anisum, L.; fennel seed, Foeniculum vulgare, Mill.'

The contrast between /y/ and its absence and that between /v/ and its absence become weakened in the following situations. Phonemically, the sequences yi, yɪ, vu, uv, ov never occur (but see 64 for an exception); vʊ occurs only initially in a few Persian loanwords; ye,

vo are never found in V– and are infrequent elsewhere; iy, ɪy, ʊv are never found in –V and are infrequent elsewhere; ey is infrequent. Phonetically, prevocalic i, ɪ, ʊ, o (/u/ is never prevocalic) and postvocalic e, o have corresponding semivocalic transitions as described in Table 4 except in sequences of any two out of i, ɪ, u, ʊ.

TABLE 4

The phonetic correlates are as shown below given that the adjacent vowel is :

	Front :	Central :	Back :
H ɪ in –V* (Similarly **H** i)	ɪy	ɪy	ɪy ~ ɪ
U ɪ in –V* (Similarly **U** i)	ɪy ~ ɪ	ɪy ~ ɪ	ɪ
H ʊ in –V*	ʊw ~ ʊ	ʊw	ʊw
U ʊ in –V*	ʊ	ʊw ~ ʊ	ʊw ~ ʊ
H o in –V	ow ~ o	ow ~ o	ow
U o in –V	o	o	ow ~ o
H, U e in V–	ye ~ e	ye ~ e	e
H o in V–	o	wo ~ o	wo
U o in V–	o	o	wo ~ o

V* is any V other than i, ɪ, u, ʊ.

Examples are : gaɪe 'please sing', pɪe 'drank (m. pl.), drink (contingent, 2nd and 3rd sg.) ', pɪa 'drank (m. sg.)', prɪa ' beloved (fem. of prɪy) ', pɪo ' drink (imperative 2nd pl.) ', b^{h}aɪo^{n} ' (oblique pl. of b^{h}ai, brother) ', dvɪtia 'second (fem. of dvɪtiy) ', č^{h}ʊe ' touched (m. pl.), touch (contingent, 2nd and 3rd sg.)', č^{h}ʊa ' touched (m. sg.) ', sʊas 'sweet smell (cf. vas, smell) ', č^{h}ʊo 'touch (imperative 2nd pl.)' **H** [č^{h}ʊwo], **U** [č^{h}ʊwo ~ č^{h}ʊo], taʊo^{n} '(oblique pl. of taʊ, father's brother)', soi ' slept (f. sg.)', sao ' anet, dill seed, Anethum sowa, Kurz. = Peucedanum graveolens, Benth.', sou^{n} ' sleep (contingent, 1st sg.)', soo+ge will

sleep (2nd pl.)' H [sowo], U [sowo ~ soo], k^hee 'rowed (m. pl.), row (contingent, 2nd and 3rd sg.)'. gae 'sang (m. pl.), sing (contingent, 2nd and 3rd sg.)', gae^n 'sing (contingent, 1st and 3rd pl.); cows (pl. of gay)', soe 'slept (m. pl.), sleep (contingent, 2nd and 3rd sg.)', k^heo 'row (imperative, 2nd pl.)', deo^n 'gods (oblique pl. of dev)', gao 'sing (imperative, 2nd pl.)', b^hao^n (oblique pl. of b^hav, rate)', $\check{c}^h$ʊo above, soo+ge above. But no semivocalic transitions in : vaqɪi 'indeed', šʊur 'good manners', sʊi 'needle'.

(72) ɪy [ɪyə̆] ≈ i before a consonant or # in Sanskrit loanwords, as : kṣətrɪy 'member of the second varṇa', b^hartiy 'of India', prɪytəm 'beloved'. Similarly with iy ≈ i, ʊv ≈ u, ey ≈ e : ǰatiyta 'communalism, caste-ridden nature', d^hrʊv 'Pole Star', varṣney '(surname)', paṇdey '(surname)'. Compare : k^heya 'rowed (m. sg.)'.

(73) e ∞ y occuring after a non-front vowel in a few items, as : ba^ne ~ bae^n '(oblique of baya, left)', dae^n 'to the right', H bəǰae 'in place of' (U bəǰaɪ, see 39). The two endings—contingent 2nd and 3rd sg. and contingent 1st and 3rd pl.—regularly alternate e ∞ y, H ≈ ve, as in : gae 'sing (contingent, 2nd and 3rd sg.)', gae^n 'sing (contingent, 1st and 3rd pl.)', soe+ga 'will sleep (2nd and 3rd sg.)', soe^n+ge 'will sleep (1st and 3rd pl.)'. There is no such alternation, say, in : gae 'sang (m. pl.)', pərae 'stranger (adj., m. sg. oblique and pl.)', gae^n 'cows'.

(74) o ∞ v occurring after a non-back vowel in the contingent and imperative 2nd plural ending, as in : k^heo 'row (contingent and imperative 2nd pl.)', gao 'sing (contingent and imperative 2nd pl.)', kərao+ge 'will have (it) done (2nd pl.).' There is no such alternation, say, in : deo^n '(oblique pl. of dev, god)', b^hao^n '(oblique pl. of b^hav, rate)', $pɪtao^n$ '(oblique pl. of pɪta, father).'

Note further the morphophonemic loss of y in i–V, ɪ–V, V–e (as in : dvɪti–a, prɪ–a, ga–e^n 'cows' above) and

of v in u–, ʊ–V, o–, V–o (as in : sʊ–as, de–o^{n}, b^{h}a–o^{n} above). There is a similar loss of y in –i, –ɪ and of v in –u, –ʊ, as in :

vɪṣə–i 'given to sensual pleasure' (cf. vɪṣəy 'subject matter; earthly object'), snehmə–i '(fem. of snehməy, full of friendly warmth)', b^{h}a–ʊk 'sentimental' (cf. b^{h}av 'sentiment').

The three-way contrast between i, ɪ, and y and that between u, ʊ, and v is not firmly established. FIRTH's proposal[22] to recognize only two units in either case was a brilliant stroke but too drastic to meet the situation in Hirdu (described in Table 5 and notes to that table).

TABLE 5

	Environments	Palatal vocoids			Labiovelar vocoids		
1	#—#	i	(ɪ)	—	u	(ʊ)	—
2	C—#	i	(ɪ)	(y)	u	(ʊ)	(v)
3	#—C	i	ɪ	—	u	ʊ	(v)
4	C—C	i	ɪ	(y)	u	ʊ	(v)
5	V—#	i	(ɪ)	y	u	—	v
6	V—C	(i)	(ɪ)	y	(u)	(ʊ)	v
7	V—V	—	ɪ	y	—	ʊ	v
8	C—V	(i)	ɪ	y	—	ʊ	v
9	#—V	(i)	(ɪ)	y	—	—	v

The double line shows the major line of the two-way contrast. When broken, it suggests that even the two-way contrast is weakened. Parentheses indicate that the occurrence is infrequent. C stands for any nonsyllabic.

22. In his Introduction to A. H. HARLEY, *Colloquial Hindustani*, London : Kegan Paul, 1944; see also T. Grahame BAILEY, *Teach yourself Hindustani*, London : English Universities Press, 1950, which is edited by J. R. FIRTH and A. H. HARLEY.

Notes to Table 5 (the number refers to the serial number of the environment) : (Note 1) **H** ɪ ʊ in the names of the letters; **U** ɪ ʊ do not occur here.

(Note 2) For **H** ɪ ʊ, see 35–7; for **U** ɪ ʊ, see 37–40; y [yə̆] and v [və̆] occur in Sanskrit loanwords; v [və̆] (≈ o) in Persian loanwords.

(Note 3) For non-phonemic [ĭ] here, see 95 below; v occurs only in Sanskrit loanwords (see 90).

(Note 4) y [yə̆] and v [və̆] occur only in Sanskrit loanwords.

(Note 5) For **U** ɪ, see 39; **H** ɪ does not occur here.

(Note 6) i, u in Persian loanwords (i ≈ (ɪ ∞ y), u ≈ (ʊ ∞ v) when in a free syllable) as : rəis 'rich', aina 'mirror', taus 'peacock' bəʔid 'far' (see 11). ɪ, ʊ in Sanskrit, Persian, and English loanwords (ɪ ∞ y, ʊ ∞ v when in a free syllable) as : daɪtv 'responsibility', dəɪta 'wife', bʰaʊk 'sentimental'; qaɪd 'leader', daɪrə 'circle', saʔɪrə 'the wandering one (Muslim female given name)'. mʊšaʔɪra 'poet's meet' (see 11, 115); saɪkɪl 'bicycle', haʊs 'house (in names of buildings)'; koɪla 'coal'. See also 87, 88 below.

(Note 7) For the placement of the tonic ɪ, ʊ are short : ''bʰa.ɪ'oⁿ, ''ta.ʊ'oⁿ above. See also 63, 65 above, 75 below.

(Note 8) i in Sanskrit and Persian loanwords (always i ≈ ɪ) as : dvɪtia 'second (fem.)'; niət 'intention', ɪnsaniət 'humanity'. The ɪ : y and ʊ : v contrasts are undeniable, as with : kɪa 'did', kya 'what ?'; sʊər 'pig', svər 'vowel, tone' (v chiefly in Sanskrit, Persian, and English loanwords). There is a good deal of alternation, however (76–86 below). For the placement of the

---

FIRTH uses i and y, u and w. Later followers have used ɪ and ʊ for his y and w, presumably because their ɪar 'friend' (my and FIRTH's /yar/) looks less bizarre than FIRTH's kwch 'some' (my and their/kʊčʰ/).

tonic ɪ, ʊ are of course short; kɪ''a, sʊ''ər above. See also 75.

(Note 9) i, ɪ in Sanskrit and Persian loanwords (i always i ≈ ɪ) as : ɪətta 'limit, quorum'; ias 'full of despair' (cf. yas 'despair'), ɪˀadət 'courtesy call on a sick person', ɪˀadə 'return' (see 11).

The difficulties in the way of not recognizing a three-way contrast are posed by the loanwords (see Notes to Table 5) and by the fact that prevocalic ɪ, ʊ (Table 5 and notes 7–9) are consistently short though not seriously in contrast with i, u and though phonetically higher in a free syllable. The gaps in the distribution, both complete and virtual, are, however, undeniable (Table 5 and notes), and so are the alternations based on them (see Notes to Table 5 and below).

The sequences ii, iɪ, ɪi, ɪɪ, uu, uʊ, ʊʊ have not been encountered except for the following Persian loanwords : vaqɪi 'indeed', šʊur 'good manners' (but see 114).

(75) ɪ before a vowel other than i, ɪ; ⊙ before i, ɪ; i elsewhere, as in : pɪa 'drank (m. sg.)', pɪuⁿ 'drink (contingent, 1st sg.)', pi 'drank (f. sg.)' and 'drink (imperative, 2nd sg.)', pina 'to drink' (corresponding to the forms gaya, gauⁿ, gai, ga, gana respectively of {ga–sing}; to gaɪe corresponds, however, piǯɪe); bʰaɪoⁿ '(oblique plural of bʰai, brother)'; sʊɪaⁿ 'needles', sʊi 'needle'; lərkɪaⁿ 'girls', lərki 'girl'. Similarly with ʊ, ʊ ∞ ⊙, u, as in : čʰʊa, čʰʊuⁿ ∞ čʰuⁿ, čʰʊi, čʰʊɪe, čʰu, čʰuna from {čʰu—touch} ; taʊoⁿ '(oblique plural of tau, father's brother)'; ǯadʊi 'magical', ǯadu 'magic'; arzʊeⁿ 'longings, wishes', arzu 'longing, wish'.

In the environment C–V, there are various possibilities in the choice between ɪ and y :

(i) ɪ (never y) as in : hɪa 'heart', dɪa 'lamp', pɪa 'beloved (fem.)', mɪaⁿ '(Muslim title of respect)', ke + zərɪe 'by means of', redɪo 'radio'; mɪˀa 'intestine'

(see 11); dvɪtia (see Note 8 to Table 5); pɪa, pɪu^{n}, etc. (see 75 above; also dɪa, lɪa, kɪa, etc. irregularly from dena 'to give', lena 'to take', kərna 'to do').

(ii) y (never ɪ) in non-initial syllables as in : kamyab 'successful', əfyun 'opium'.

(iii) any of the alternations below.

(76) y, **U** ≃ ɪ occurring in C–V in initial syllables, as in : kya 'what ?', tyohar 'festival', d^{h}yan 'attention, meditation', syahi 'ink', pyaz 'onion', myan 'scabbard', zyadə 'in excess', ḍyuṭi 'duty, tax'.

(77) **H** y, **HU**, **U** əy, **U** ≃ ɪ occurring in C–V in initial syllables, as in : syana 'wise', xyal 'opinion, (name of a genre of Hindustani classical music)', byalis 'forty-two'.

(78) ɪ, **HU** ≈ y, occurring in C–V in Persian loanwords, as : lɪaqət 'ability, wisdom', rɪayət 'concession', ke + dərmɪan 'between', vahɪat 'pointless', hošɪar 'alert', pɪadə 'pawn (in chess)'.

(79) **H** y, **U** ɪ, occurring in C–V in non-initial syllables in Sanskrit loanwords, as : əyodd^{h}ya 'Ayodhyā, the capital of Rāma', səttyagrəh 'satyāgraha'.

(80) ɪ ∼ y, **U** ≃ y, occurring in C–V in non-initial syllables in Persian loanwords, as : dʊnya 'world', fəryad 'appeal for succour or justice', dərya 'river', bʊnyad 'base'.

(81) y ≈ ɪ, **U** ≃ ɪ, occurring in C–V in non-initial syllables in English loanwords, as in : kərfyu 'curfew'.

(82) ɪ ∼ y, **U** ≃ ɪ, occurring in C–V, as in : hət^{h}ɪar 'weapon', g^{h}əṛɪal 'gavial, Gangetic crocodile, Gaviəlis gangeticus, L.', č^{h}ɪalis 'forty-six', č^{h}ɪasəṭ^{h} 'sixty-six', č^{h}ɪasi 'eighty-six', č^{h}ɪanve 'ninety-nine'.

With v and ʊ in C–V, the possibilities are slightly less varied and are as follows :

(i) ʊ, never v, as in : sʊa 'large needle', sʊər 'pig', pəčʰʊa 'west wind', sʊas 'sweet smell', mʊərrɪx 'chronicler', dʊal 'leather thong', sʊɪŋg + məšin 'sewing machine';[23] čʰʊa, etc. in 75 above (as also : hʊa, etc. irregularly from hona 'to be, to become, to happen'); dʊʔa 'prayer, praise', mʊʔərrɪq 'sudorific (sb.)' (see 11); pʰʊhara 'spray' (see 32).

(ii) any of the alternations below.

(83) ⊙ ≈ v ≈ əv, U ⊙ occurring in C–V in initial syllables in Persian loanwords, as : xvab 'dream', xvahɪš 'longing'. (Inserting v in Persian loanwords under 83, 84 is a spelling pronunciation; Urdu speakers always drop it.)

(84) ⊙ ≈ v, U ⊙ occurring in C–V in Sanskrit and Persian loanwords, as : vɪšvəmbʰər '(male given name)'; ʊpvas 'religious fast'; tənxvah 'salary', dərxvast 'application'. Compare : koṭa 'quota'.

(85) vV* ≈ ʊ (where V* is ɪ, ə) occurring in C–V in Sanskrit and English loanwords, as : dvɪtia 'second (fem.)', dvɪvedi '(surname)', əddvɪtiy 'unmatched', svəbʰav 'nature', svədesi 'Swadeshi', tvəča 'skin', dʰvənɪ 'sound; overtone of meaning'; kvɪnin 'quinine' (also kʊnɛn). Thus svəraǰy 'self-government' is rendered homonymous with sʊraǰy 'good government' (both sʊraǰ).

(86) v ≈ ʊ occurring in C–V; as in : dvar 'door, entrance', svikar 'acceptance', vɪddvan 'learned'; šəlvar 'kind of loose trousers', əfvah 'rumour', ǰəlvə 'display of oneself', həlva '(name of sweet preparation)'; kvʌlɪti 'good quality'; kvar 'the seventh month of the Hindu calendar, Aśvina', pʊrva 'East wind', təlva 'sole (of the foot)'.

---

23. This is of course a spelling pronunciation for the English /'soɪŋg+mə'šin/; the pronunciations /sɪvɪŋg/ and /syʊɪŋg/ are also not unknown.

In V–C where C is not followed by a vowel, ɪ and yɔ alternate in some items.

(87) ɪ, **HU** ≈ yə occurring in V–C where C is not followed by V in Persian and English loanwords, as: ɣaɪb 'out of sight', laɪq fit (for), capable', nʊmaɪš 'exhibition, fair', ša'ɪr 'poet' (see 11); ṭaɪm 'time (as measured by clock)'.

(88) **H** yə, **HU** ɪ, **U** yə occurring in V–C where C is not followed by V, as in: gʰayəl 'wonded', kayər 'cowardly', koyəl 'koel, Eudynamis scolapaceus, L. and other songbirds of the same genus', ǯəyənti 'jubilee', pənčayət 'panchayat, the village or caste council', narayəṇ '(male given name)', šayəd 'perhaps', kɪfàyət 'profit', ayəndə 'future (adj.), in future (adv.)'.

Finally, the position of y and v is insecure even in #–V, V–V not to speak of V–C, C–V, #–C, C–#, C–C, witness the following alternations. Persian and English loanwords are unaffected (but see 91).

(89) **H** y, **HU** y ~ ǯ, **U** ǯ occurring in #–V, V–V in some Sanskrit loanwords, as: yəmʊna '(female given name, name of river)', yʊddʰ 'war', yʊg 'era', əyoddʰya 'Ayodhyā, the capital of Rāma'. Similarly **H** y, **HU**, **U** əǯ in C–#, C–C, as in: sury ~ surəǯ 'sun'. Compare: ǯ in ǯəntər + məntər 'magic', ǯɔ 'barley, Skt. yava'; y in yəh 'this' and in the loanwords yɔn 'vaginal, sexual', yar 'friend', yorəp 'Europe'.

(90) **H** v, **HU** v ~ b, **U** b occurring in #–V, #–C, V–C, C–V in some Sanskrit loanwords, as: vən 'forest', vɪdya 'learning', vɛdy 'Ayurvedic physician', vyakrəṇ 'grammar', vrət 'religious fast', vrɪkṣ 'tree', kərtəvy 'duty', pərvət 'mountain', somvar 'Monday', səvⁿvət ≈ səmvət ~ səmbət 'year of the Vikram era'. Similarly **H** v, **HU**, **U** əb in C–#, C–C, as in: purv ~ purəb 'east'. Compare: b, **H** ≃ (b ~ v) in: brəǯ 'Braj', bəsənt 'spring (season)', bibʰəts 'repulsive', baṇ 'arrow', brɪhət 'big', bɪndʊ

'point'; b in: bɪhar 'Bihar', bag^h 'tiger'; v in vəh 'that, he, she, it' and in the loanwords veššya 'prostitute', vəzir 'vizier', voṭ 'vote'.

(91) **H** v, **HU** v ≈ b, **U** v occurring in V–V, C–V, and intervocalically in gemination in a few items, as: səvera 'morning', bava 'old man, holy man', bivi 'wife', nəvab 'Nawab', ǯəvab 'reply', ɪkyanve 'ninety-one', nəvve 'ninety'.

The semivowels in Sanskrit loanwords in other positions remain unaffected.

One problem involving the semivowel y has been postponed to § 1.2.7 (see 105 below).

1.2.7 The Problem of geminates and other clusters.

The contrast between a single consonant and the corresponding geminate is quite common in $V\text{–}(^h)V$, especially $\breve{V}\text{–}(^h)V$. (Compare also 28, 30, 62–65 above.)

Elsewhere there are difficulties. Sequences of the type $/hC_1C_1/$ are seen in 29 above.

(92) **H** $C_1C_1(^h)[\breve{ə}] \simeq C_1C_1(^h)ə$, **HU**, **U** $C_1(^h)$ in V–# in Sanskrit loanwords, as: pɪtt 'bile, one of three humours in Ayurvedic medicine', $yudd^h$ 'war', $budd^h$ 'Buddha', udatt 'acute accent; sublime (adj.)'; $dukk^h$ 'pain' (cf. 32 above).

(93) C_1, **U** $\simeq C_1C_1[\breve{ə}]$ in V–# in Persian loanwords, as: šəkk 'doubt', zɪdd 'determination', həqq 'truth, God; privilege', xətt 'letter, document'. Compare: off '(exclamation of surprise)'; d^həmm ∼ d^həm 'onomatope for a thud); xəttɪ + ɣulami 'document of slavery'.

In sequences of the type $VC^*(^h)C^{**}$ where C^* is any consonant or v and C^{**} is l, r, h, y, v in syllable final position, phonetically C^* is fortis and long and C^{**} is folowed by [$\breve{ə}$]. (/ṛ/ never occurs in the C^* position.)

mɪtr 'friend', pɔtr 'son's son', sɔty 'truth', tətʰy 'fact', dɛty '(kind of evil being in Hindu myth), daitya' ǯaḍy 'lethargy', kərtəvy 'duty', tətv 'principle'; qətl, 'slaughter', ɪtr 'attar', fəth 'victory', həšv 'stuffing, padding'.

In sequences that are similar to these except in that C** is followed by a vowel and does not include h, phonetically there are two possibilities.

(i) C*(ʰ) ends the syllable and is followed by [ĭ] before y or [ə̆] before l, r, v. (/ḍ/ and /ŋ/ never occur in this position.)

badloⁿ 'clouds (oblique plural)', əsli 'genuine', ʊtra 'descended', ɪtroⁿ 'attars (oblique pl.)', pɪtroⁿ 'ancestors (oblique pl.)', səmraṭ 'emperor', gumrah 'one who has lost one's way', ʊpyog 'use', dɔrya 'river', əvyɔv 'limb', əfyun 'opium', bʰəgvan 'God', manvi 'human', vɪšvas 'belief'; isvi 'Christian (era)', pəṛʰvana 'to have (it) read (by someone)'.

(ii) C*(ʰ) is ambisyllabic, fortis, and long. (/ŋ/ and /ṛ/ never occur in this position.)

dʰɔnaḍʰyoⁿ 'wealthy persons (oblique pl.)'. (Compare sɔŋgya 'term, noun', see 19 above).

With the exception noted below this is predictable and involves no contrast. The exception is Sanskrit loanwords with intervocalic sequences of the following types where they correspond to a cluster in Sanskrit: stop other than ḍ followed by C** or ʰC**, nasal other than ŋ followed by y or v, sulcate or r or l or v followed by y.

(94) C_1, H ≃ C_1C_1 in V-(ʰ)C**V, especially in V̆-(ʰ)C**V, in the sequences just described, as in: mɪttroⁿ 'friends (oblique pl.)', puttri 'daughter', mattra 'measure', vɪppləv 'disturbance', səttyagrəh 'satyāgraha', aggya 'command', vɪddvan 'learned', sənnyas 'renunciation

of this world, sannyāsa', ənnyay 'injustice', ənnveṣəṇ 'research', veššya 'prostitute', nɪrryat 'export', kəllyaṇ 'welfare', kavvyo^n 'long poems (oblique pl.)' (cf. 67).

In ḍ(^{h}) followed by y or v only the second possibility obtains. In all other cases not covered by 94, only the first possibility obtains. Examples have already been given earlier.[24]

Note the morphophonemic simplification of a geminate elsewhere; as in: mɪttro^n 'friends (oblique pl.)', but mɪtr 'friend'; əddvɪtiy 'unmatched', but dvɪtiy 'second'; krɨttɪka 'the Pleiades, the third lunar mansion', but kartɪk 'the eighth month of the Hindu calendar'; vɪkkrɨt 'distorted', but sənskrɨt 'Sanskrit', krɨtɪ 'action'; prəsənn 'glad, pleased', but prəsənta (more carefully: prəsənntạ) 'gladness, pleasure'. (Compare: pɪtər 'ancestor', ɪtr 'attar', with the respective oblique plurals pɪtro^n and ɪtro^n; and pətt^{h}ər 'rock' with the oblique plural pətt^{h}əro^n and the derivative pət^{h}rila 'rocky'.)

Note the treatment of sequences in Sanskrit loanwords corresponding in Sanskrit to clusters of the type of hC* (where C* is m, n, ṇ, r, l, y, v) and to the sequence hṛ. Initially: hrəsv 'short', hrɨdəy 'heart' (see 96, 97, 98). Intervocalically: bahyaŋg 'outward aspect', ǯɪhva 'tongue', čɪhno^n 'marks (oblique pl.)' (see 52, 106). Finally: čɪhn 'mark', brəhm 'Brahma, the cosmic principle' (hm, hn, hṇ subject to 52, 100, 106; hr, hl, hv subject to 52, 106; hy unaltered).

Hirdu has on the whole an imperfect tolerance of initial and final sequences of two or more nonsyllabics.

---

24. On this problem from a wider and different angle, see P. B. PANDIT, Borrowing—a study of linguistic expression of social distance, *Vidyā, Journal of the Gujarat University* (Oct. 1961) 1-24; idem, Sanskritic clusters and caste dialects, *Indian ling.* 24 (1963) 70-80.

This leads to the use of supporting vowels [ĭ], [ə̆] and /ɪ/, /ʊ/, /ə/.

Initial clusters with y or v as the last member have been dealt with in 76–8, 83–6 above. The only initial cluster seen at all in Persian loanwords is /xv/ (but see 83, 84). In the formulas that follow C is any nonsyllabic.

(95) H əs ≃ s [ĭs ~ ə̆s], HU, U ɪs ≈ sə in #–C* where C* is p, t, ṭ, k, m, n, as in: spəṣṭ 'explicit', spənǯ 'sponge', sprɨšy 'not Untouchable', sp^hurtɪ 'agility', stər 'layer', stri 'woman', st^han 'place', st^hai 'permanent (adj.); the opening of a song in Hindustani music (sb.)', sṭešən 'railway station', skul 'school', skru 'screw', sk^həlɪt 'that which has slipped', snan 'bath'. (smərəṇ 'memory' has sʊmɪrəṇ sporadically by the side of the regular smərəṇ, ɪsmərəṇ, səmrəṇ. Note that *səməraṇ is not possible; see § 1.2.8.) Similarly with the initial cluster šm, as in: šməšan 'Hindu crematorium'.

(96) rɪ ≈ ɪrɪ ≈ ɪr in #C(C)–C*, where C* is a heterosyllabic C, as in: krɪkĕṭ 'cricket (game)', kripa 'favour, mercy' (see 57 above). Similarly ri ≈ ɪri ≈ ɪr, as in pritɪ 'love', pritəm '(given name)'; rʊ ≈ ʊrʊ ≈ ʊr, as in trʊṭɪ 'defect'.

(97) rɪ ≈ ɪrɪ in #C(C)– not followed by a heterosyllabic C, as in krɪa 'verb; action', krit + yʊg 'the Golden Age (in Hindu myth)'. Similarly ri ≈ ɪri in #C(C)–, as in: šri 'Mr.'; ru ≈ əru ≈ ʊru in #C(C)–, as in: fruṭ 'fruit as dessert'.

(98) r ≈ ər in #C(C)–V* where V* is any vowel other than i, ɪ, u, ʊ as in: praṇ 'breath of life', ṭrəŋk 'trunk, box', b^hrəm 'illusion', šrəm 'labour', vrət 'fast'. (Note that prəkaš 'light' will yield not *pərəkaš but pərkaš; see § 1.2.8.) The use of additional familiar forms like tɪras, dʊroh, pɪrkaš respectively for tras 'fear, trouble', droh 'betrayal', prəkaš 'light' is best regarded as sporadic. The same is true of the simplification seen

in historical doublets liket g^hɪna ~ g^hʊna ~ g^hɪn ~ g^hʊn and kɪšən ~ kɪsən by the side of g^hriṇa 'aversion' and kriṣṇ '(Hindu male given name), Krishna'.

(99) l ≈ ɪl in #C(C)–V, as in pleg 'bubonic plague', glas 'glass tumbler', mlan 'languid', sleṭ 'slate', šlok 'stanza in a classical Sanskrit metre'. The initial clusters sl, šl are alternatively treated as under 95. Where the vowel after l is u, ʊ the alternants are l ≈ ɪl ≈ ʊl, as in: fluṭ 'flute'.

Note that /s/ in #—C* is phonetically [ĭs ~ ə̆s] where the extra-short vowel will not count in the placement of the tonic: **H** ≃ /stri/ [ĭ"stri ~ ə̆"stri] 'woman', but "ɪs.tri, "əs.tri, "sət. ri 'woman', "ɪstri 'flat iron'. Note also the syllable divisions. /ṛ/ in # — V is [ə̆ṛ]; /ṇ/ in # — V is [ə̆ṛ̥ ≃ ṇ] (see § 1.2.3 for both).

The phonetics of the final clusters with l, r, h, y, or v as the last member has already been described. See also 41 above which alone is applicable if the last member is identical with the second last (but see also 28, 92, 93), or is homorganic with a preceding nasal (but see 19), or is y or v (see also 89, 90), or is a stop with a sulcate preceding. If the last member is h, 49 (and not 41) is applicable. If the second-last member is h, 100 (and not 41) is applicable (see also 51, 52, 106; əhC has no final [ə̆]). In all other cases, both 41 and 100 are applicable (compare 95).

(100) C_2[ə̆] ≈ əC_2 in C_1—(^{h})# occurring subject to the reservations in the foregoing paragraph, as in: mɪtr 'friend', gəṇtəntr 'republic', rʊgṇ 'patient (sb.)', sukṣm 'minute (adj.)', b^həsm 'holy ashes', rəkt 'blood', ǯənm 'birth', $murk^h$ 'foolish', $tirt^h$ 'holy water', brahm + mʊhurt 'early morning', čɪhn 'mark'; vəqt 'time', qətl 'slaughter', ɪtr 'attar', lɔfz 'word', mʊft 'free of charge', vəzn 'weight', ʔʊmr 'age', ərz 'request', gərm 'hot', bərf 'snow, ice', hərǯ 'loss, injury', xətm 'completed'; bɔks 'box', kəmyunɪzm 'communism', bolṭ 'bolt', korṭ 'law court'; ʊṛd 'black gram'.

To some of the clusters coming under 100, the following (101, 102) are also applicable in Persian and English loanwords—if the last vowel in the word is ɪ or ʊ and is followed by s, l, r, h.

(101) C_2 [ə̆] ≈ ɪC_2 occurring after ɪs, ɪl, ɪr, ɪh in conditions noted above, as in : qɪsm 'kind', ʔɪlm 'knowledge', sɪrf 'only', mɪhr 'sun, love'; fɪlm 'film, cinema', sɪlk 'silk'.

(102) C_2 [ə̆] ≈ ʊC_2 occurring after ʊs, ʊl, ʊr, ʊh in conditions noted above, as in : hʊsn 'youth', mʊlk 'territory', ʔʊrf 'by-name, alias', kʊhl 'kohl (for darkening the eyelids)'.

Certain nonsyllabic sequences deserve a separate mention whether constituting or being contained in initial, final, or medial clusters. They undergo certain simplifications. The simplification of geminates has already been described.

(103) **H** ≃ kṣ, **HU** (č) $č^h$, **U** (k)k^h (geminates intervocalically; otherwise simplified) occurring in Sanskrit loanwords, as in : kṣətrɪy 'member of the second varṇa', dəkṣɪṇ 'south', ləkṣməṇ '(male given name)', kṣəma 'forgiveness', ekakṣ 'one-eyed'. (/kṣ/is also subject to 2, 41, 100. Note, however, that the substiution of k^h for ṣ (when ṣ is not preceded by k) is sporadic and is seen in historical doublets like b^haṣa and b^hak^ha 'language'; cf. 2.)

(104) **H** šč, **HU**, **U** (č) $č^h$ (geminate intervocalically, otherwise simplified), occurring in Sanskrit loanwords, as : pəščɪm 'West', hərɪščəndr '(male given name)' pʊnščəli 'prostitute'.

(105) y ≈ ⊙ after č, $č^h$, ǯ, $ǯ^h$, ṣ, š, mostly in Sanskrit loanwords (which are also subject to 94), as in : əččyʊt '(male given name)', vačy 'voice (in grammar)', ǯyamɪti 'geometry', trɪǯya 'radius', svəraǯy 'Swaraj', šɪṣy 'disciple (of a guru)', šyam '(male given name)', veššya

'prostitute' (≈ veššа ≈ veša), vɛšy 'member of the third varṇa', əvəšy 'certainly'; ɪmšyorɛns 'insurance' (hypercorrect for ɪnšʊrɛns ~ ɪnšorɛns); ǯyada 'in excess' (from zyadə by applying 6, 43); čʰyalis 'forty-six' (see 82). (Compare also 84.)

(106) $hC_1* \approx C_1*h$, H ≃ $hC*_1C*_1$ in V—V (where C* is m, n, ṇ, r, l, v and the whole sequence corresponds to a cluster or to hṛ in Sanskrit) in Sanskrit loanwords, as : brahməṇ 'Brahman, member of the first varṇa', sʊhrid 'friend', ǯɪhva 'tongue', čɪhnoⁿ 'marks (oblique pl.)'. Similarly : $hC*_1 \approx C*_1h$ in V—#, as : brəhm 'Brahma, the cosmic principle', čɪhn 'mark', hy; H ≃ hyy in V—V, as : bahyaŋg 'outward aspect'. Contrast : səhvas 'living with, company', səhyatri 'fellow traveller', kʊhl 'kohl (for darkening eyelids)'.

Hirdu is much more tolerant when it comes to medial clusters. Some of the problems have already been dealt with (see 29, 94, 103–106 above). The crucial problem, however, is the status of [ə̆].

The morphophonemic loss of /ə/ in a medial free syllable in an atonic position is quite regular and yields many clusters.

ʊtərna 'to descend', but ʊtra 'descended'; ləhər 'wave', but ləhreⁿ 'waves'; pʊstək 'book', but pʊstkeⁿ 'books' ["pʊstə̆.kẽ]; ləṛəkpən 'boyhood' but ləṛka 'boy'; smərəṇ 'memory', but səmrəṇ (95); prəkaš 'light' but pərkaš (98).

The principal exception is that ə is always retained in C—V. (See also 115 below.)

H rʊpya and rʊpye ["rʊpĭ.ye] 'rupee(s)', but HU, U rʊpɛa and rʊpəe [rʊ.pə"(y)e] respectively; sənšəy 'doubt' but sənšəi 'doubter'.

The rise of a medial cluster around a morpheme cut is also quite common.

ərtʰ ≃ ərtʰə followed by vad yields ərtʰvad 'fallacy' ["ərtʰə̆.vad] and not *[ər.tʰə"vad] (the second /ə/ will drop out); mɪtr ≃ mɪttrə followed by ta yields mɪtrta 'friendship' ["mɪtrə̆.ta] and not *["mɪt.rə'ta].

Medial privative and permutational contrasts involving /ə/ and nonsyllabics are of course not ruled out.

Compare sʊər 'pig' with sʊr 'tune', akəsmɪk 'sudden, unexpected' with ləkṣmi 'Lakshmi, goddess of wealth'; šrəm (≈ šərəm) 'labour' with šərm (≈ šərəm) 'shame'; prədeš (≈ pərdeš) 'region' with pərdeš 'foreign land'; əməl 'without blemish', ʔəməl 'power, intoxication' with əml (≈ əməl) 'acid'.

There are also phonologic limits on unrelieved clusters. [ə̆] is commonly inserted at the syllable boundary and must be distinguished from phonemic /ə/.

[pʊstə̆kẽ], [ərtʰə̆vad] above; ǰyotsna 'moonlight' ["ǰyotə̆.sna], prəsənnta 'pleasure' [prə"sənnə̆.ta] (by the side of the less careful prəsənta), dʰurtta 'roguery' ["dʰurtə̆.ta], məttta 'intoxication (with power, etc.)' ["məttə̆.ta]. (The insertion of /+/ at the syllable boundary after [ə̆] does not seem to be a convincing solution.)

Our rules for predicting syllable boundaries may then have to be suitably amplified so that the insertion of [ə̆] or [ĭ] can be predicted with ease. Also sequences of the type nnt, rtt, ttt will have to be admissible. In some cases no [ə̆] or [ĭ] is inserted.

məntri 'minister, secretary' ["mən.tri], sənnyas 'renunciation of this world, sannyāsa' [sən"nyas] (∼/sənyas/ [sənĭ"yas]).

Perhaps there are phonologic limits on the insertion of [ə̆] too, so that some cases of /ə/ in a medial free syllable in an atonic position are tolerated.

əsmərəṇ (in 95 above); pʊnščəli 'prostitute'; santvəna 'consolation, assurance'. (*əsmrəṇ, *pʊnščli

*santvna are not tolerated while apparently daršnɪk 'philosophical' (by the side of dəršən 'philosophy')' sənskrit 'Sanskrit', b^hərtsna 'reproach' are tolerated at least in **H**≃.)

gəḍərɪa '(name of a shepherd caste)' (the cluster ḍr is being avoided).

The possibility cannot be ruled out that there are cases of permutational contrasts arising out of the insertion of [ə̆] and that in some environments [ə̆] and [ə] are variants ([ə̆], **H**≃ ə[ə]).

Further study has to be made before the revised rules of syllable division, those of the insertion of [ə̆], [ɪ̆], and those of the insertion (or retention, as the case may be) of /ə/ can be framed. It is unlikely that the rules of the placement of the tonic will be disturbed.[25]

One final word about [ə̆]. When the scope of certain intonational contours ends in a consonant—especially the scope of hesitation and suspension contours, [ə̆] is inserted under certain conditions. In **U** this particular variety of [ə̆] is phonetically rather high and retracted. Also [ə̆] is optionally inserted in **H** ≃ VC—#, as in, $əd^h$ɪk 'in excess'.

There are sporadic cases of the simplification of a cluster with the so-called compensatory lengthening of the preceding vowel.

(107) **H** $\bar{V}C_1$, **HU** $\bar{V}C_1 \sim \breve{V}C_1C_1$, **U** $\breve{V}C_1C_1$ in a few items as: mak^hən ~ $məkk^h$ən 'butter', gid^h ~ $gɪdd^h$ 'vulture'. Compare also similar other historical doublets (**H** first, **HU** both, **U** second) : $buṛ^h$a ~ $buḍḍ^h$a 'old man', $gəṛ^h$a ~ $gəḍḍ^h$a 'pit, deprssion', šəkər ~ šəkkər 'sugar', $rək^h$a ~ $rəkk^h$a 'kəpt', $uṭ^h$a ~ $uṭṭ^h$a 'arose', big^ha ~

25. The whole problem of ə and consonant sequences is of course reminiscent of the similar problem in French that Maurice GRAMMONT tried to resolve with only partial success by his "three consonant law".

bɪgha '(a measure of land)' (34 is a also applicable), ni^{n}bu ~ nɪmbu 'lime (citrus fruit)', b^{h}a^{n}g ~ bhəŋg 'bhang, narcotic produce of Cannabis sativa, L.', pəčis ~ pəččis 'twenty-five', sətais ~ səttais 'twenty-seven' (similarly sətavən 'fifty-seven', sətasi 'eighty-seven', sətanve 'ninety-seven'), əṭ^{h}ais ~ əṭṭ^{h}ais 'twenty-eight' (similarly əṭ^{h}avən 'fifty-eight', əṭ^{h}asi 'eighty-eight', əṭ^{h}anve 'eighty-nine').

1. 2. 8. The Problem of atonic short vowels.

The short vowels /ɪ ʊ ə/ are liable to vary among themselves and with their absence when in an atonic position. Occasionally this extends to other positions also (see 108, 109, 111, 113). The familiar variant in some cases under 113–117 has almost ousted the received variant.

(108) ə, HU ≈ ⊙ in C_1—C_2 # in Persian loanwords (⊙ hypercorrect), as : ɣələt 'wrong', mərəz 'disease', ɣəzəl 'ghazal (type of lyric)', ʔərəq 'sweat, liquor, essence'. (Compare səhər, rəhəm in 51 above.)

(109) ɪ, HU ≈ ⊙ in #—sC in a few items, as : ɪstɪʔmal 'use' (≈stemal, cf. 45 above), ɪstifa 'resignation', ɪspat 'steel'.

(110) ə, HU ≈ ⊙ in #C_1—C_2 in an atonic position in a few items, as : nəhana 'to bathe', bəhɪšti 'water carrier', ǯəhaǯ 'ship', məharaǯ 'sir' (all with C_2 changing into /^{h}/); kəroṛ 'crore, 10^7' (also kəṛoṛ), gəvah 'witness', səvari 'passenger'. Compare : H əṛ^{h}ai, U ḍ^{h}ai 'two and half'.

(111) ə ∞ ⊙ in C_1—C_2 where C_2 ends a non-final syllable in a few items, as : ʊnəttis 'twenty-nine' (∞ ʊntis), ɪkəttis 'thirty-one', mʊhəmməd '(male given name)' (∞ mʊhməd), mʊhərrɪr 'amanuensis'. (Compare məhətv, məhəbbət in 51).

(112) ʊ ≃ ɪ in C_1—C_2 in an atonic position in a few items, as : ke + sɪpʊrd 'entrusted to'. Similarly : ɪ ≃ ʊ, as in : gʊroh 'group, crowd'.

(113) ə ≈ ɪ in C_1—C_2 in a position other than a medial atonic free syllable, as : gəya 'went', pəčhəttər 'seventy-five' (the first ə), məšin 'machine', nəmək 'salt' (the first ə). Similarly : ɪ ≈ ə, as in : ke + ʔɪlavə 'apart from', nɪvaṛ 'broad cotton tape', əlɪf 'alif, the first letter of the Persian alphabet', pəṇḍɪt '(title for a learned Brahman)', məndɪr 'temple', purṇɪma 'full moon', dəkṣɪṇ 'south', čʰɪhəttər 'seventy-six'; **H, U** ɪ ≈ ə, **HU** ə as in : bəhɪn 'sister', sahɪb 'sahib; sir', vapɪs '(come, give) back', mɔsɪm 'weather, season', bahɪr 'out', dʊlhɪn 'bride', sʊhagɪn 'married woman', səyyɪd '(name of a clan, Muslim title of respect)', xalɪs 'pure; sincere'; ə ≈ ʊ, as in : səfed 'white', yorəp 'Europe'; ʊ ≈ ə, as in : hʊzur '(title of respect)', sʊluk 'good treatment'; **H, HU** ə, **U** ʊ∼ə, as in : qʊsur 'error in duty', fʊzul 'superfluous', mʊsʊlman 'Muslim' (second ʊ), sʊbut 'evidence', mʊʔaf 'pardoned', sʊhulət 'concession, convenience'; ə, **U** ≃ ʊ, as in : mʊɣʊl '(ethnic name)' (second ʊ).

(114) ə, **U** ≃ ɪ before i in Persian loanwords, as in : vaqɪi 'indeed'. Similarly : ə, **U** ≃ ʊ before u, as in : šʊur 'good manners'.

(115) ʊ ≈ (ə ∞ ⊙) in a medial atonic free syllable, in some items : yəmʊna '(female given name, name of a river)', ǯʊmʊʔə 'Friday', ustʊrə 'razor' (also ≈ əstʊra). Similarly ɪ ≈ (ə ∞ ⊙), as in : kəvɪta 'poem', ɪʔtɪraz 'objection', ɪntɪzar 'waiting', xalɪsə 'the king's own property'; ə ≈ (ɪ ∞ ⊙), as in : mʊšaʔəra 'poets' meet', mʊamələ 'matter, affairs', mʊhavərə 'idiom', pərvərdəgar 'God as protector'; ʊ ≈ ɪ ∞ ə ∞ ⊙, as in : mʊskʊrana 'to smile' (second ʊ); ə ∞ ⊙, as in : əsəpʰəl 'unsuccessful' (second ə), kripəya 'kindly, please', vɪḍəmbəna 'mockery' (second ə), məsələn 'for example', mɔləvi 'scholar', ʊləma 'the learned', ke+rubəru 'face to face with', zyadəti 'presumptiousness', bɪstəra 'bedding', sadəgi 'simplicity', karigəri 'craftsmanship'. ǯɔhəri 'jeweller', mɪhnəti 'industrious', bevəquf 'foolish',

əspətal 'hospital', məsʔələ 'problem' (after applying 11 and 43). Compare ə (never lost) in: senapətɪ 'army commander'.

Finally a few alternations involving /ə/ adjacent to a vowel may be taken up, all in Persian loanwords.

(116) a, **U** ≃ əə in Persian loanwords, as: təʔəʒʒʊb 'surprise', təʔəllʊq 'relation', məʔən 'suddenly, at once'. Similarly a, **U** ≃ aə, as in: ʒəmaʔət 'class, group', vɪdaʔə 'fare-well'; a, **U** ≃ əa, as in: səʔadət 'well-being', məasɪr 'good deeds; good omens', məal 'effect', mʊamələ 'matter, affairs', ke + mʊafɪk 'according to', mʊʔaf 'pardoned'. (Some of these require prior application of 11 and 113.)

(117) ⊙, **U** ≃ ə in Vʔ—# in Persian loanwords, as: bədiʔə 'strange', vəziʔə 'bad, evil', šʊruʔə 'that which has begun'.

1. 2. 9. A Selection of sporadic correspondences.

The following list is of course only a representative sample of the kinds of correspondences that one might expect to encounter. (Compare also the sets 22, 32, 38, 47 (the last example), 54, 63, 65, 73, 74, 91, 98, 103, 107, 109–112 above; see also § 1. 2. 3, last paragraph.)

kəp^{h}, **HU** ≈, **U** ≈ kəf 'phlegm, one of the three humours in Ayurvedic medicine' (similarly p^{h}əl 'fruit'); nəməskar ≈ nəməškar 'greeting with folded hands (also interjection)'; prəsad, **HU** ≈ pəršad 'offering to God, grace of God', **H** murtɪ, **HU**, **U** murət 'idol'; veššya ≈ vɛššya 'prostitute' (see also 105); raʒa, **U** ≃ raʒə 'king'; narayəṇ, **U** nərayən '(Hindu male given name)'; **H** brahməṇ, **U** brəhmən 'Brahman, member of the first varṇa'; yəmʊna, ʒəmna, **U** ≃ ʒəmən '(name of river)' (but only the first two for the Hindu female given name); **H** ʊṣa, **HU** uša 'dawn, (Hindu female given name)'; sɪŋh (see further 21), sɪ^{n}h ∼ sɪn^{h} ∼ sɪn^{h}a 'Singh, Sinha (in names, respectively from Western and Eastern areas)'.

The use of **H** a, **U** ə [ă] finally in proper names like Ashoka Hotel, Gupta, Mishra, Sinha is a pronunciation influenced by the Romanizations of these names.

pil ~ fil 'elephant'; **H** ǯurab, **U** ǯʊrrab 'stocking'; **H** pərvah, **U** pərva 'regard, concern' (similarly lapərva(h)i 'negligence'); ɪstifa, **U** ≃ ɪstɪʔfa 'resignation (from a post)';[26] **H** nəgaṛa, **U** nəqqarə '(kind of drum)'; gʊbbaṛa ~ gʊbbara, **U** ≃ ɣʊbarə 'bomb, balloon' (cf. ɣʊbar 'cloud of dust'); **H** gʊmməṭ, **HU** gʊmməǯ, **U** gʊmbəz ~ gʊmbəd 'dome'; **H** gəfar, **U** ɣəffar 'clement (epithet of God, Muslim male given name)'; **U** ≃ dəfʔə 'time, occasion' (ending in 'ain and subject to 11, 43 yielding dəfa; in phrases like do + dəfa 'two times', dusri + dəfa 'for the second time', there is also found **HU** dəfe; cf. also **U** ≃ dəfʔə 'section of a law', **U** ≃ dəfʔə 'who has been pushed away', used in the phrase dəfa + ho + ǯao 'be gone!, away with you!'); **H** čoga, **U** čʊɣa '(kind of robe for men)'; hɪndʊstan, **U** ≃ hɪndostan 'India; Upper India' (23 is also applicable); šəhzadə ~ šahzadə 'prince, son of a king'; ʊmmid, **U** ʊmmid ≈ ʊmed 'hope'; **H, HU** mɪad, **U** miʔad ≈ mɪʔad 'period, duration'; dukan ~ dʊkan 'shop'; bečara, **U** bečara ≈ bɪčara 'pitiable, poor'; **H** nehru, **HU**, **U** nəhru '(surname)' (cf. nəhr 'canal' under 52 and čɪhra 'face' under 45); **H** k^hʊsro, **HU** xʊsrɔ ≈ xʊsru, **U** xʊsrɔ 'famous (as title)'; rastə, **HU** ≈ rəstə 'road'; bazar, **HU** ≈ bəzar 'market, bazaar'; **H** səhulɪət, **U** sʊhulət ~ səhulət 'concession, convenience'.

rɪporṭ ≈ rəpəṭ 'report'; ɪsṭɛṭɪks ≃ stɛṭɪsṭɪks 'statistics' (sic!); brɪṭɛn ~ brɪṭɛ̆n, **U** bərtanɪa 'Britain'.

H čɪṭk^həni, **U** čɪṭxəni ~ čəṭxəni 'bolt'; p^hɪr, **HU** ≈, **U** ≈ fɪr 'again'; **H** pənčanve, **HU** pəčanve, **U** pɪčanve

26. **U** ≃ distinguishes between ɪstɪʔfa 'asking to be excused; resignation (from a post)', ɪstɪfa 'acceptance; being accepted', ɪstifa 'fully carrying out of an obligation' (spelt respectively *isti'fa, iṣṭifā,* and *istīfā*). **H, HU, U** confuse between them in various ways.

'ninety–five'; ɪkyavən ~ ɪkavən 'fifty–one' (similarly ɪkyasi 'eighty-one', ɪkyanve 'ninety-one'); nɪnanve ~ nɪnnanve ~ nɪnyanve 'ninety-nine'; byanve ~ banve 'ninety-two'; čay ≈ ča 'tea'; sʊhavna ~ sʊhana 'pleasant' (cf. sʊhana 'to please'); **H** tɪrsəṭ^{h}, **U** tresəṭ^{h} 'sixty-three' (subject to 98; similarly tɪrpən' 'fifty-three'); nəhi^{n} ∞ nəi^{n} 'not; no (opp. yes)'; dʊpəhr ~ dopəhr 'afternoon'; kəṭ^{h}p^{h}oṛva ~ kəṭ^{h}p^{h}ʊṛva 'woodpecker'; **H** sɪlgana, **U** sʊlgana 'to kindle' (similarly : sʊləgna 'to be kindled'); **H** č^{h}ɪpna, **U** č^{h}ʊpna 'to hide (intrans.); ɪkkis~ɪkkɪs 'twenty-one' (similarly with the numerals twenty-two to forty-eight except tis 'thirty'); **H** či^{n}ṭi, **HU** či^{n}vṭi, **U** čyu^{n}ṭi 'ant'; **H** ek, **HU**, **U** ek ≈ ɪk 'one' (ɪk also in Urdu verse); **H** ǰa + kər, **HU** ǰa + kər ~ ǰa + kər + ke ~ ǰa + ke, **U** ǰa + ke 'having gone' (similarly with bol+kər 'having talked' and so on; except **H, HU, U** kər + ke 'having done'); **H** ʊs + pər, **HU** ʊs + pər ~ ʊs + pe ≈ ʊs + pɛ, **U** ʊs + pe 'on him, her, it, that'; **H** yo^{n}, **HU, U** yu^{n} 'this way' (similarly tyo^{n} 'that way', kyo^{n} 'why?'); **H** to, **HU, U** to ≈ tɔ 'then (correlative of 'if')'; **H** kašmir, **U** kəšmir 'Kashmir'; **H** məraṭ^{h}i, **HU, U** mərhəṭi (50) 'Marathi'; kona ≈ kɔna 'corner'; tyohar ~ tyɔhar 'religious festival.'

It will be noticed that numerals have prominently figured in the display of alternations. The ones from eleven to eighteen are especially troublesome : **H** gyarəh ≈ gyar^{h}ɛ ≈ gyar^{h}e (like 53), **HU, U** gyarəh ≈ gyar^{h}a ≈ gyara (like 54) 'eleven'; **H** gyarəho^{n}, **HU, U** gyarəho^{n} ≈ gyar^{h}ao^{n} ≈ gyarao^{n} (like 54) 'eleven (oblique pl.)'; **H** gyarəhva^{n} ≈ gyar^{h}ɛva^{n} (like 52), **U** gyarəva^{n} ∞ gyarva^{n} (cf. 115) 'eleventh' (similarly with twelve to eighteen; eleven has the additional peculiarity of **U** ≈ ɪgyarəh, etc., cf. 109; seventeen has the additional peculiarity of **H** sətra ~ səttra, etc., as in 94, though it is not a Sanskrit loanword; fifteen and seventeen are subject to the observation under footnote 17 when subject to 52, 53, 54, Rule ix).

Apart from the numerals, one other important group of forms exhibiting sporadic correspondences is the proximal and the distal/unmarked pronouns.

H yəh [yəh] ≈ ye ≈ ʒe, **HU** yəh [yɛ̆h] ~ ye, **U** yeh ≃ yɪh [yĕh] 'this' (cf. the emphatic form **H, HU, U** yəhi, **U** ≃ yɪhi [yĕhi] 'this alone'); **H, HU** ye, **U** ye ≃ yɪh 'these' (cf. the emphatic form, **H, HU, U** yehi 'these alone'); **H** vəh [vəh], **HU** yəh [vɛ̆h] ~ vo, **U** vo ≃ vʊh [vŏh] 'that; he, she, it' (cf. the emphatic form **H, HU, U** vəhi, **U** ≃ vʊhi [vŏhi] 'that alone; etc.'); **H** ve, **HU** ve ~ vo, **U** vo 'those, they' (the emphatic forms will be **H** vehi, **HU** vehi ~ vohi, **U** vohi, **U** ≃ vʊhi). The adverbs yəha^n 'here', yəha^n 'there' are always in that form, yha^n ~ hya^n and vha^n ~ hva^n being distinctly substandard.

Occasionally these sporadic correspondences are historically doublets.

H sɪr, **U** sər 'head' (respectively Indo-Aryan and Persian); **H** -st^{h}an, **U** -stan '-land (especially in names)' (respectively Indo-Aryan and Persian); **H** ʒira, **U** ≃ zirə 'cumminseed' (respectively Indo-Aryan and Persian); əspətal ~ həspətal ~ hʌspɪṭəl 'hospital' (respectively Portuguese and English).

The following case does not amount to a doublet but is nevertheless of interest : **H** karryəvahi; **U** karrəvaˀi, **HU** karvai 'putting into action, procedure'.

1. 3. A Consolidated picture.

The inventory presented § 1.1 stands revised (as shown in Table 6) with 27 consonants (28 + 13 including r^h − 14 including r^h = 27), 4 nonsyllabic vocoids (3 + 1 = 4), 13 syllabic vocoids (8 + 5 = 13), and 1 coarticulation (1 = 1).

q ˀ ɣ are accompanied by faucal constriction especially in U≃. ṇ is [r̥̃ ≃ ṇ] (with partial nasalization of tautosyllabic vocoids) when neither initial nor adjacent to another retroflex consonant nor before a syllable final

y or v. ṛ is [r̃] (with full nasalization of tautosyllabic vocoids) when tautosyllabic with /^{n}/. /^{h}/ (aspiration)

TABLE 6

Consonants (C) :							
Stops :	p	t	ṭ	č	k	q	ʔ
	b	d	ḍ	ǯ	g		
Nasals :	m	n	ṇ		ŋ		
Sulcates :		s	ṣ	š			
		z		ž			
Fissurates :	f				x		
					ɣ		
Lateral :		l					
Tremulants :		r	ṛ				
Nonsyllabic vocoids (W) :		h	^{h} (aspiration)				
	y						v

Syllabic vocoids (vowels, V) :	i				u
		ɪ	ɨ	ʊ	
	e	ĕ			o
	ɛ		ə		ɔ
			a		A

(The vowels ɪ, ɨ, ʊ, ĕ, ə are printed inside a box.)

Coarticulation : ^{n} (nasalization)

can occur after any consonant that is neither a fricative (sulcate or fissurate) nor q ʔ ṇ ŋ; it is extra-lenis syllable-finally after a voiced consonant in HU∞. (The aspirates in the original inventory can now be eliminated.) The vowels in the box are short and neutral; the remaining vowels are long and peripheral. The rules for the placement of the tonic and the subtonic are applied accordingly. ɪ ʊ are [ĕ ŏ] in U and U ≃ under certain circumstances. əhʊ is [ɔ̆h] in H∞, U when followed by C. ə is [ă] finally

in **U**. ə is [ɛ̆] in **HU** finally and before h with certain limitations. ə is [ɛ̆] in **U** before h with certain limitations.

It is highly unlikely that **H** ṇ ṣ ɨ on the one hand **U** q ʔ ž x ɣ on the other will coexist in the same idiolect. Some idiolects can get along without any or all of the following consonants: q ʔ ṣ š z ž f x ɣ. Possibly some use [ṇ], [ṛ] (either or both of them) only as variants of /n/, /ḍ/ respectively. Possibly some use [ʰ], [y], [v] (any or all of them) only as variants of /h/, /ɪ/, /ʊ/ respectively. The introduction of ʔ serves to resolve some cases of syllable-division within a phonologic word that would otherwise be unpredictable and distinctive and to account for some distributions of [ĕ ŏ].

Some idiolects can get along without any or all of the following vowels: ɨ ĕ ʌ. Possibly some use [ɛ] and [ɔ] only as positional correlates of the sequences /əy/ and /əy/ (or even of /əɪ/, /əʊ/) respectively. Possibly some idiolects are hyperdifferentiated and the variants [ĕ] of /ɪ/, [ŏ] of /ʊ/, and [ɛ̆ ɔ̆ ă] of /ə/ (any or all of them) assume a phonemic status. Some idiolects use [ĭ], [ə̆] sparingly and /ɪ ə/ more freely; others use [ĭ], [ə̆] freely and /ɪ ə/ more sparingly. For the latter class of idiolects, the rules of syllabic division may have to be amplified.

The incipient polarization between [V̌N] (where N is a nasal homorganic with the following consonant) and /V̄ⁿ/ is possibly complete in some idiolects, so that [VN] can be interpreted as /Vⁿ/ (leading to the elimination of /ŋ/ possibly) with a suitable revision of the tonicity rules.

In sum, little or no amplification of the original phonemic inventory is called for in the case of some idiolects. Such idiolects are probably closer to, if not identical with, lowbrow casual Hirdu (Fig. 2, no. 4). In this simplified picture h, ⁿ, and what there is left of y, v are

probably better integrated systemically with the rest of the nonsyllabic inventory. (Some forms marked by ≈ as inelegant or familiar are of course particularly contributory towards a restricted inventory.)

1.4. Some wider implications of the present study

A sociolinguistic note should perhaps be added. In contrast to some other modern Indian languages, Marathi for example, the cleavage in Hirdu between formalized and casual (Fig. 2, between 1 and 2 on the one hand and 3 and 4 on the other) is very sharp (so as to allow us to speak of vertical diglossia), while the cleavage between standard (highbrow and middlebrow) and substandard (lowbrow) is not so sharp. This observation is probably valid not only for word phonology but also for other portions of the language system and is probably to be correlated with the socio-cultural background.

This three-layered situation—standard Hindu, substandard Hirdu, and 'lower languages'—may represent a somewhat attenuated continuation of the earlier trinity of Classical Sanskrit, Prakrits-and-Apabhraṁśa, and the lower languages of pre-New-Indo-Aryan times in which the first was the more or less exclusive vehicle of the Great Tradition and its renovation.[27] Perhaps the Hirdu writer's penchant for classicalizing his language (whether in the direction of Sanskritization or Persianization) which deepens the diglottic cleavage between 'higher' and 'lower' is to be explained in the light of this historical background.

The complexity of the phonology of contemporary Hindi–Urdu may be traced to an extensive dialect mixture

27. Cf. J. Bloch, *L'Indo-aryen*, Paris : Adrien-Maisonneuve, 1934, pp. 20–1, 322.

through borrowing and substratum influence : a typical case is (107) under § 1.2.7.

On the practical plane, this study, it is hoped, will serve to correct the overly simplified and **U**-oriented picture of Hirdu that many Western linguists and pedagogues carry with them.[28] It will also serve to lend more body to the rather abstract picture drawn in the introduction (§§ 0.1–2) and to justify the use of the complex apparatus set up there (§ 0.3). For the purposes of teaching Hindi-Urdu to persons outside India and Pakistan, a comprehensive description will perhaps be of special interest. This is also true even if only a passive or receptive control of a number of speech varieties is aimed at. More specifically, it may be suggested that (a) the pattern congruity between [ḍ] and [ṛ] on the one hand and [ṇ] and [r̥̃] on the other should perhaps be brought out; (b) /ʰ/ and /ⁿ/ are pedagogically more useful symbols than /h/ or /ʿ/ for the first or /˜/ for the second; (c) for any teaching level but the very rudimentary, an overdifferentiated vowel symbolism is probably desirable, say on the lines suggested in Table 7 (with, of

TABLE 7

Neutral :			Peripheral :		
i [ɪ]	ɨ [ɨ]	u [ʊ]	i: [i]		u: [u]
e [ĕ]		o [ŏ]	e: [e]		o: [o]
ɛ [ɛ̆]	ə [ə]	ɔ [ɔ̆]	ɛ: [ɛ]		ɔ: [ɔ]
	a [ă]			a: [a]	ɒ: [ʌ]

Cluster-supporting vowels : ⁱ [ĭ], ᵊ [ə̆]

ɛ ɔ a; i: u: e: o: are tense. The rest are lax.

---

28. For instance the monophthongal rendering of the nuclei of *bail* 'bull' and *aur* 'and' is overemphasized. After all [bɛe̯l], [bəe̯l], even [bəi̯l] is as good Hirdu as [bɛl]; [bɛe̯l] is used even by some "Urdu" speakers.

course, suitable eliminations where feasible); (d) a foreign speaker should probably avoid trying to be over-smart and say, /ɪsṭešən/ or /səṭešən/ for /sṭešən/ 'railway station' (95–102 so far as they affect Sanskrit and English loanwords) ("Hindi" speakers at any rate do not relish this).

The vowel phone inventory given in Table 7 is pleasingly symmetrical ([ɨ] and [ᴀ] can be easily eliminated for many practical purposes) and probably very handy in sorting out the complex allophonics of **H**, **HU**, and **U** vowels. It will be a mistake, however, to hastily promote it to the status of an overall phonemic inventory. It is possible to work the insights of a generative handling of phonology into a descriptive framework that is more widely accessible without being simplistic.

That a language system has both a core and margins has long been suspected. (By core I do not mean merely an overlap of a set of related systems; a core in the present sense can presumably be discovered even if we are considering a single language system. HOCKETT has used the term in both senses. The term margin has been used in the way BAZELL has used it.) Can we now express the hope of making use of this distinction for a more adequate presentation of linguistic facts—the hope of refining upon, as it were, the separation of some rules by small print seen in many a traditional grammar? The present study has attempted to do something in this direction.

The reader's attention may be drawn to the consistent use of the rules of predicting the placement of the tonic for the purposes of resolving many problems of non-unique phonemicizations. Many would agree in principle that, if any feature is to be considered non-phonemic, it must clearly be shown to be either predictable and so redundant on the one hand or paraphonologic on the

other hand. But at the same time many have neglected to apply the principle to non-phonemic word accent. It may be further suggested that if we work out rules for predicting the tonic within a word for a number of widely separated modern Indo-Aryan languages,[29] we may perhaps obtain a more reliable clue to the genetic subgrouping. Mutual interference of the type illustrated in 107 above is much less likely to obscure the basic splits in the case of the "subconscious" area of redundant features than it would in other cases.

Establishing the phonologic syllable is not strictly necessary for the operation of the tonic placement rules. (For example, a light syllabic may be defined simply as a short vowel that is final or followed by another vowel with or without a single consonant in between.) But

---

29. For instance: I have earlier analysed Marathi word accent (The Phonology and morphology of Marathi, Cornell University Ph. D. thesis, 1958, § 1.3.4). Dr. L. M. KHUBCHANDANI has worked out rules for Sindhi word accent (at the Kurukshetra Seminar in December 1963; see his Stress in Sindhi, *Indian linguistics* 27 (T. P. Meenakshisundaram volume, 1966, in the press) or the chapter on Stress in his forthcoming monograph *Sound structures of Sindhi and Hindi: A Contrastive study*, Poona: Deccan College). Dr. S K. CHATTERJI's work on Bengali suggests that word accent tends to fall on the initial syllable: the status of his Sanskrit loanword exceptions is, however, not quite clear (see his A Brief Sketch of Bengali Phonetics, *BSOS* 2 : 1 (1921) 1–25, also reprinted by International Phonetic Association, London, 1921, §§ 58, 60). My own work on Kashmiri suggests that the tonic is always the first syllable of a phonologic word (see A. R. KELKAR and Pran Nath TRISAL, Kashmiri Word Phonology: a first sketch, *Anthropological Linguistics* 6 : 1 (Jan. 1964) 13–22, § 1 and note 7). In Panjabi the placement of the tonic is not wholly predictable: pə′ṛna 'to read', pərvá 'consideration'; tə̄ṛka 'dawn', təṛkā 'burst mustard in hot oil (imperative)'; pìaḷ 'shareholder', bɪàr 'Bihar' (′, -, ‵ respectively standing for prominence accompanied by high falling, tonally unmarked, and low falling-rising pitch).

certainly the phonologic syllable makes it neater; it is also useful in defining conditions for allophonic variants and morphophonemic alternants. (Consider the use of the terms *tautosyllabic*, *heterosyllabic*, *syllable-final*, *free syllable* in this study.) I have emphasized this because the syllable has been much maligned—and deserves, I think, a reconsideration.

Quite apart from any of these applications, however, I shall feel amply rewarded if I succeed in inducing others to feel and enjoy in some measure the chaotic beauty of this complex phonologic system.

ADDENDA AND CORRIGENDA

In spite of great care certain errors and omissions have crept in. Only the less obvious ones are noted below. The reader is urged to enter them in the body of the book before reading the book. When the lines are counted from the bottom of the page, the letter b is added to the number. See also the note on p. xii.

page	line(s)	
8	10b	*after* West Pakistan; *insert* it is the secondary official language of Andhra Pradesh;
21	2b	*delete comma*
26	2	*for* b^{h}ai *read* b^{h}a.i
27	10–11	*for* or (in case there is no coda) *read* preceded) by a short vowel and in all other cases
29	8b	*for this line read* U rʊpəya [rʊ.pə"ya]
31	11	*for* ṇ (21) *read* ṇ (20)
33	13	*for* kafi *read* kʌfi
34	2b	*for* NN*C* : re^{n}gana *read* VN*C* : rə^{n}gana
34	1b	*for* reŋgin *read* rəŋgin
35	3	*for* V^{n}P≈VP *read* V̆^{n}P≈V̆P
35	12	*for* casaes *read* cases
35	8b	*for* [Ṽŋg ∞ Ṽŋ] *read* [Ṽ^{ŋ}g ∞ Ṽŋ]
36	6	*for this line read* (19) Vŋg [V̨ŋŋ] before a nasal; Vŋg [V̨ŋg$^{ə̆}$ ∞ V̨ŋŋ]
36	7	*for* [Vŋg ∞ Vŋŋ] *read* [V̨ŋg ∞ V̨ŋŋ]
37	8	*after* sulcate, *insert* r,
37	9–10	*for* pəṇḍIt 'Hindu scholar', *read* pəṇḍɪt '(title for a learned Brahman)',
37	5b	*for* zemi^{n} *read* zəmi^{n}

42	8b	*for* [ḥ] *read* [h̥]
43	20	*for* cf. fn. 14 end *read* cf. also fn. 17 end
43	24	*for this line read* slow tempo), as in: mə^{n}ǯ^{h}d^{h}ar 'midstream' (∞ mə^{n}ǯd^{h}ar),
43	25	*for* keṭ^{h}p^{h}ʊṛva *read* kəṭ^{h}p^{h}ʊṛva
44	14	*after* 32 *insert* ; cf. also 92 below
55	4	*for* vɛšnəv *read* vɛṣṇəv
56	3b	*for* sao *read* soa
57	13	*delete* b^{h}artiy 'of India',
57	15	*after* nature', *insert* b^{h}artiy 'of India',
57	16	*for* paṇdey *read* paṇḍey
60	16	*after* uʊ *insert* ʊu
64	2b	*after* position *insert* and where C* and C** are not identical
70	4b–3b	*for* **HU**, **U** rʊpɛ̣a and *read* **HU** rʊpɪa and rʊpɪe, **U** rʊpəya ∞ rʊpɛa and
75	16b–15b	*read* the sets 22, 32 (the last example), 38, 48 (the last example), 54, 63, 65, 66 (the last example), 73, 74, 77, 91, 95 (the parenthesized example), 98, 103, 107, 109–112
76	2b	*in the three italicized words add a macron above the first* a *and put the umlaut sign below the second* t
77	15b	*for* festival.' *read* festival'; **H** rʊpya, **HU** rʊpɪa, **U** rʊpəya ∞ rʊpɛa 'rupee' (and the respective plurals rʊpye, rʊpɪe, rʊpəe; cf. the sets 60, 82 and page 70).
78	8b	*after* presented *insert* in